Integration of Vulnerable Road Users Behavior into a Virtual Test Environment for Highly Automated Mobility Systems

René Degen, M.Sc.
Cologne University of Applied Sciences, Germany,
Uppsala University, Sweden.

Alexander Tauber, M.Sc.
Cologne University of Applied Sciences, Germany.

Marcus Irmer, M.Sc.
Cologne University of Applied Sciences, Germany,
Uppsala University, Sweden.

Alexander Nüßgen, M.Sc.
Cologne University of Applied Sciences, Germany,
Uppsala University, Sweden.

Florian Klein, M.Sc.
HHVISION GbR, Cologne, Germany.

Dr. -Ing. Christian Schyr
AVL Deutschland GmbH, Karlsruhe, Germany.

Prof. Dr. Eng. Mats Leijon
Uppsala University, Sweden.

Prof. Dr. rer. nat. Margot Ruschitzka
Cologne University of Applied Sciences, Germany.

Abstract:

Human road users are the most vulnerable participants in urban traffic. This is due on the one hand to their spatial proximity to other road users, and on the other hand to the individuality of their behavior and decision-making processes. For this reason, they play a crucial key role for the development of highly automated vehicle systems - especially for urban environments.

This thesis describes an approach to integrate real human traffic behavior into the approval and testing process of highly automated vehicle systems. It provides a safe and valid way to test critical traffic scenarios between vehicles and pedestrians. Basically, two different methodologies for the metrological detection of human movements are analyzed and experimentally investigated for their suitability for this use case. Besides the general functionality, plausibility and real-time capability are further investigation criteria. The thesis concludes with the integration of the propagated solution into a test bed for highly automated vehicle systems using a representative traffic scenario.

1. Introduction

Simulation methodologies play an important role in the development of modern mobility solutions. Thereby, mobility is often defined by technical vehicle solutions. In this paper the detection of vulnerable road users (VRU) behaviour is investigated. The aim is to capture the biomechanical behaviour in a highly realistic way and forward it to a virtual test environment for highly automated vehicle systems. This Co-Simulation environment consist of a virtual reality environment which represents urban traffic reality, an automated virtual vehicle and a VRU avatar, which is controlled by a sufficient model. The structure is shown in Figure 1.

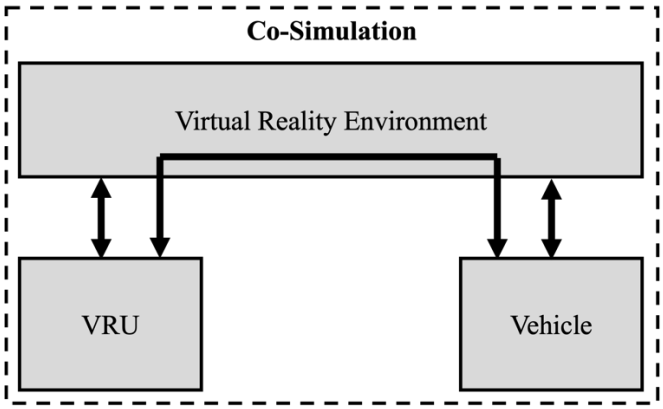

Figure 1: Overall test field structure

2. VRU Model

The aim of this chapter is to find a methodology to describe human movement behavior in a sufficient and efficient way. Therefore, two different types of measurement hardware in combination with software processing are described and investigated, before an advocated solution is chosen.

Bases for the rating of a sufficient solution is the anatomy of the musculoskeletal system. In general, the human body is divided into the body parts trunk (truncus) and two pairs of limbs (extremities), the arms (membrum superius) and legs (membrum inferius). The trunk is subdivided into the head (caput), neck (collum), and trunk (truncus or torso). The trunk further consists of the chest (thorax), abdomen (abdomen), and pelvis (pelvis). An arm includes the upper arm (brachium), forearm (antebrachium), and hand (manus). A leg includes the thigh (femur), lower leg (crus), and foot (pes). [1]

It is important, that the methodology captures the movement in a highly realistic way.

2.1. Marker-assisted Motion Capture Model of the VRU

Marker-assisted motion capture systems are based on the detection of markers that are attached to the subject's body. For this purpose, cameras are used that emit light in the infrared spectrum into the room and pick up reflective markers through a camera sensor without an infrared filter. [2]

The cameras can have any position and orientation to each other, whereby they build an epipolar geometry. Accordingly, it is necessary that at least two cameras always capture one marker in order to be able to reconstruct the depth information.

The detection volume is a result of the dimensions of the measurement setup, as well as the number of cameras and their resolution. A schematical setup is shown in Figure 2.

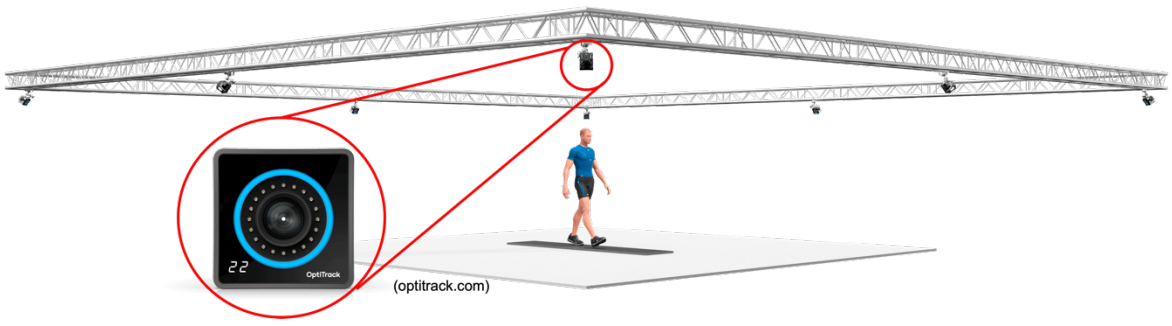

Figure 2: Schematical setup of a marker-based motion capture hardware

To prepare a Person-under-Test (PuT) for the measurement, it has to be equipped with markers at predefined positions. For this propose, the PuT wears a suit, which carriers the markers. Figure 3 shows the way, how the markers were positioned for an optimal joint identification and detection.

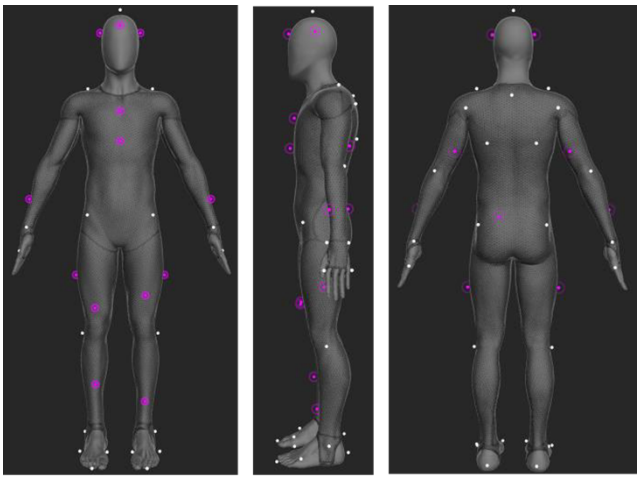

Figure 3: Marker positioning setup

The skeleton is determined by the position of the markers in the detection volume [2]. In the frame of this work, a predefined algorithm to determine the skeleton is used, which results in a 21 body joint skeleton. The result of the marker-assisted skeleton is shown in Figure 4.

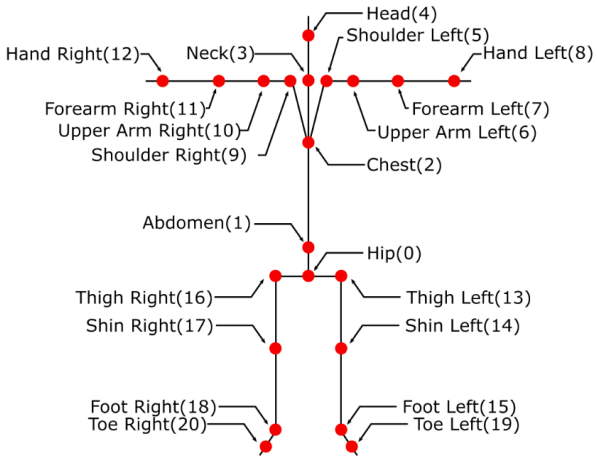

Figure 4: Schematical skeleton of the marker-assisted VRU-model

2.2. Marker-less Motion Capture Model of the VRU

Marker-less motion capture systems correspond in their function to marker-assisted systems. The difference is that the subject does not have to be prepared with markers before the recording process. These systems mostly emit light in the infrared spectrum in the form of lasers, structured light patterns or modulated light into space. A depth image of the area is determined via measurement of the time of flight in the case of lasers or recording of the distorted light pattern. [3]

For this purpose, the Kinect for Xbox 360 from Microsoft is used. Here the depth image is created by emitting a well-known dot pattern into the detection space. By analyzing the deviation of the dots from the well-known pattern, depth is determined. The depth image is then evaluated pixel by pixel to identify individual parts of the subject's body. A random forest classifier – a pretrained artificial intelligence algorithm - is used for this purpose. [4]

The progression steps of the marker-less skeleton determination are shown in Figure 5.

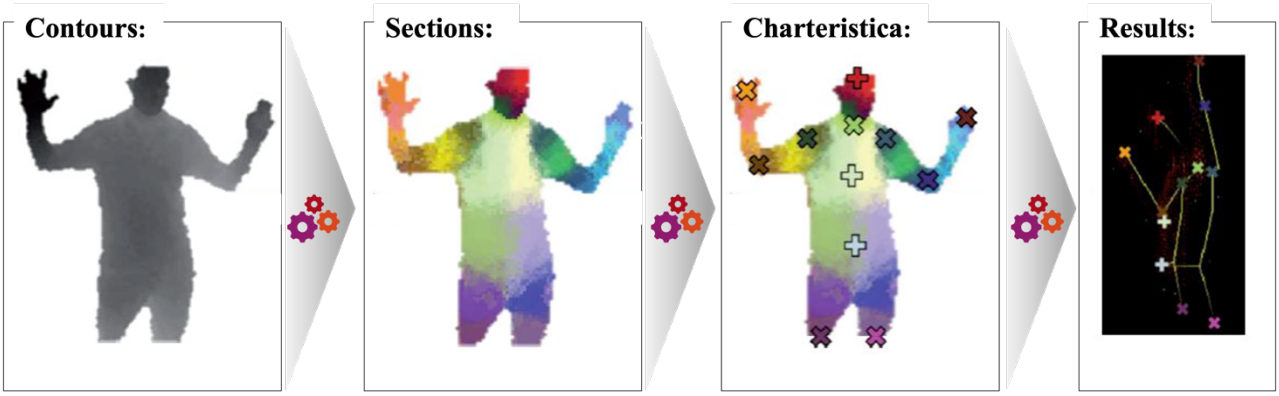

Figure 5: Progression steps of the skeleton determination [4]

To increase the detection volume and quality, and to enable a comparability to the marker-assisted tracking with different detection perspectives, a system consisting of two Microsoft Kinect cameras is used.

The two Kinects can be arbitrary positioned and orientated in space. Both cameras are connected in one network to address one synthetizing progression algorithm. Each Kinect camera determines position and rotation data for each body joint based on the individual position and orientation of the camera. For calibration this data is forwarded to the synthetizing algorithm. This algorithm has to calculate the displacement and torsion of the two Kinect cameras to each other and determines the influence to the resulting body skeleton. For this purpose, the coordinate system of the first Kinect is defined as the world coordinate system. The coordinate system of the second Kinect must be transformed into this. In [5] two methods are presented for this using the successor, the Kinect for Xbox One. On the one hand linear regression and on the other hand a genetic algorithm. These two methods are adapted on the Kinect for Xbox 360 and expanded to solve in three dimensions in the following.

For linear regression as well as the genetic algorithm, a set $n = 200$ corresponding frames per camera are defined as input data. Each frame contains the data of one entire skeleton at this specific time of one camera. After collecting the input frames the torsion from the second to the first Kinect between each body joint in each corresponding camera frame is calculated and averaged over all joints in the skeleton. This rotation information is then used to rotate the position data of the second Kinect into the orientation of the first Kinect. Based on this, the deviation in position of each joint in corresponding frames is calculated and averaged over all body joints. This information about torsion and deviation is then used to calculate the mean squared error of each pair of frames.

For the linear regression, the inverse of the mean squared error is used as weight. Thus, the estimated deviation and torsion between the two Kinects is calculated.

The genetic algorithm uses the previously described input data as starting population. These 200 individuals are then evolved over 600 generations. The selection is rank based and the recombination is uniform. Per generation the best 25% of the population are used for recombination. In addition, the top 10% of the population of each generation are passed on to the next generation unchanged as the elite. A threshold value of 0.5 is chosen for the crossing probability of a gene. If the associated gene-dependent random variable falls below this value, a crossover occurs. In addition, a mutation takes place in each gene by addition or subtraction of a value within a defined range of -5 to +5. Fitness is calculated via transforming the skeleton data of the second Kinect into the position and orientation of the first Kinect and calculating the mean squared error. In addition, the frames are weighted the same as in the linear regression. Figure 6 and Figure 7 show the genetic algorithm over 50 runs with the same starting population. Figure 6 shows the convergence of the genetic algorithm used. The fitness of the best individual with the best fitness per generation is shown relative to the individual of the last generation with the best fitness of the corresponding run. The red line represents the average over all runs and the grey area the range of all runs. Figure 7 shows the distribution of the relative deviations of the results of all runs in relation to the result with the best fitness.

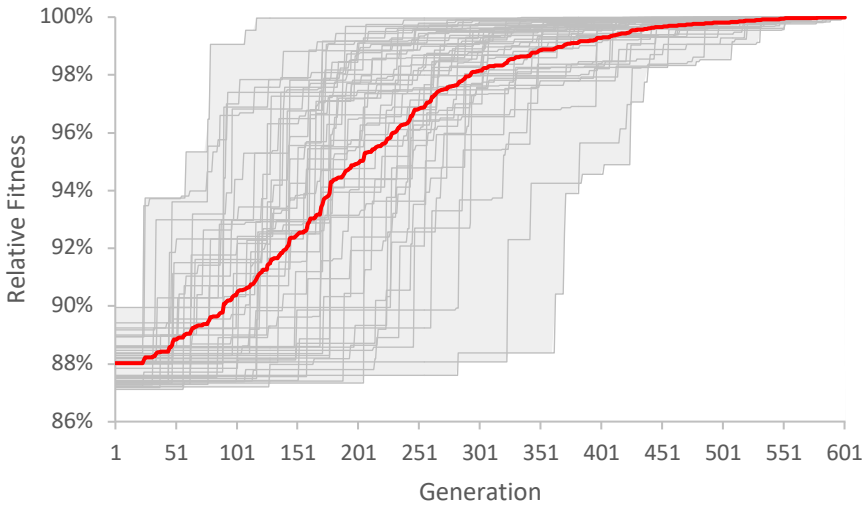

Figure 6: Convergence of the genetic algorithm used over 50 runs

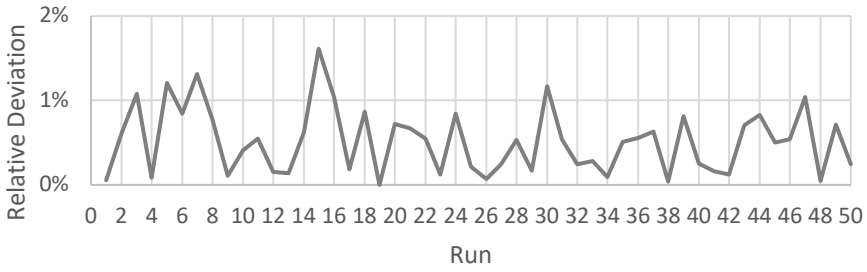

Figure 7: Deviation of the genetic algorithm used over 50 runs

The data show that the genetic algorithm over the observed calibrations always delivers better results than the linear regression.

The resulting skeleton architecture of the marker-less VRU model is shown in Figure 8.

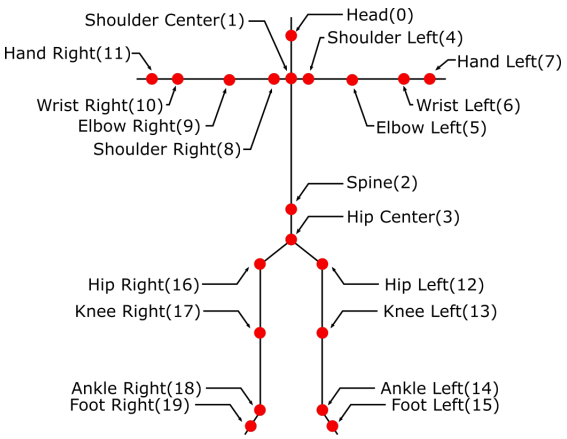

Figure 8: Schematical skeleton of the marker-less VRU-model

It is important to notice, that this model uses 20 body joints to describe the biomechanical behavior of the captured PuT.

3. Investigation and Results

In the frame of this chapter, the different motion capturing methodologies were compared in context of accuracy and level of detail. For the investigation basis, a person is captured in a static A-pose. This pose allows the shade less view of all measurement setups to the investigated person. For the PuT a male person with an approximate body height of 185 cm is chosen.

The resulting skeletons of the different motion capture algorithms (blue: marker-assisted; orange: marker-less) were shown in Figure 9 next to the PuT in A-pose.

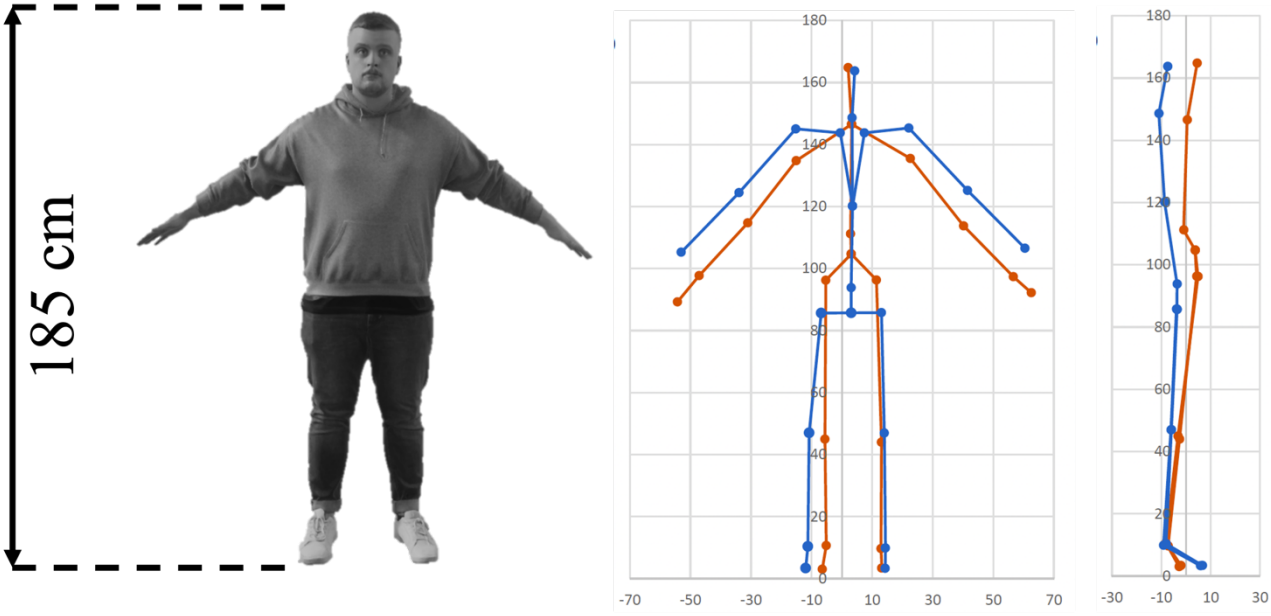

Figure 9: Resulting skeletons of the different motion capture algorithms (blue: marker-assisted; orange: marker-less) in frontal and lateral view next to the PuT in A-pose

The comparison of the skeletons shows that the toe base, ankle, and knee joints, as well as the joints of the cervical spine and the head, are approximately the same height in both systems. In the side view the connection vector between ankle joints and toe base joints in the orange skeleton is shorter than in blue one. In the marker-less recorded skeleton, these correspond to about 8 and 9 cm, respectively, and in the marker-assisted skeleton to about 16 cm, which is close to reality. Significant differences can be seen in the area of the pelvis. For example, the marker-less system estimates the position

of the pelvis to be approx. 20 cm higher than the marker-assisted system and is thus located in the abdominal region. In addition, the hip joints drop by about 8 cm compared to the pelvis. In the blue skeleton, all three joints are at the same height. The spinal column is subdivided into the three parts of the spine in a simplified manner. It consists of joints for the cervical spine, the thoracic spine as well as the lumbar spine and the sacroiliac joint, the connection of the spine with the pelvis. This creates a natural course of the spine in the side view. In the orange skeleton the spine consists of two joints and the sacroiliac joint. In the side view, there is also a clear kinking of the connecting vector between the lumbar spine and the sacroiliac joint. Furthermore, in the upper extremities there is no dedicated joint for the sternoclavicular joints. These are combined with the shoulder joints as one joint. This results in a lower position of the shoulder joints. On the other hand, there are separate joint points for the hands. However, this does not represent a joint with rotational degrees of freedom, which occurs in the passive human musculoskeletal system. The same rotation data is output as for the carpal joint. The same applies to the metatarsophalangeal joint in relation to the ankle joint.

Additionally, the rotation of the forearms around their longitudinal axis is a special challenge of optical motion detection. The cross-section of the forearm changes only insignificantly during an internal or external rotation in a two-dimensional camera image. This rotation is transmitted directly to the hand via the carpal joint. Accordingly, the elbow joint is decisive for this rotation. Figure 10 shows an external rotation followed by an internal rotation of both arms for two separate recordings in both systems. Both sequences are recorded with the palms open in a T-pose. The palms initially point toward the floor, are subsequently rotated toward the subject's line of vision, and rotated back to the starting pose.

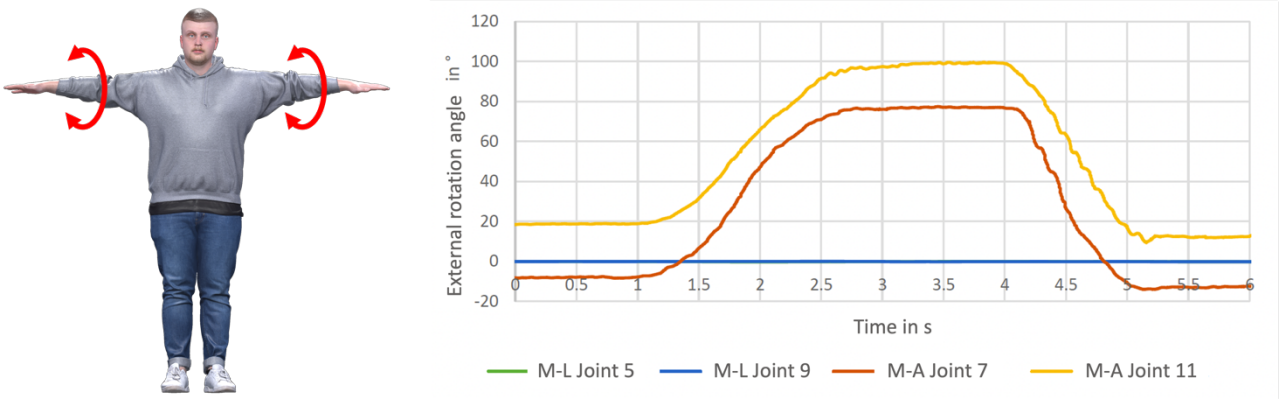

Figure 10: Left: Rotation of the forearms around their longitudinal axis in the T-pose;
Right: Results of the elbow joints for the marker-less (M-L) and the marker-assisted (M-A) skeleton determination

The marker-assisted system detects and maps the rotation of the forearm, whereas the marker-less detection system detects no rotation neither left nor right.

4. Implementation into the Virtual Test Field

Next to the biomechanical correctness, the processing time between the different systems is important. In special for testing of critical traffic situations, the real-time capability is essential. The validity of the studies depends, among other points, on brief moments in which decisions are made - both by the pedestrian and by the automated vehicle functionality.

Therefore, finally the marker-assisted motion capture model of the VRU is exemplary integrated into the virtual reality environment. The different process steps are shown in Figure 11.

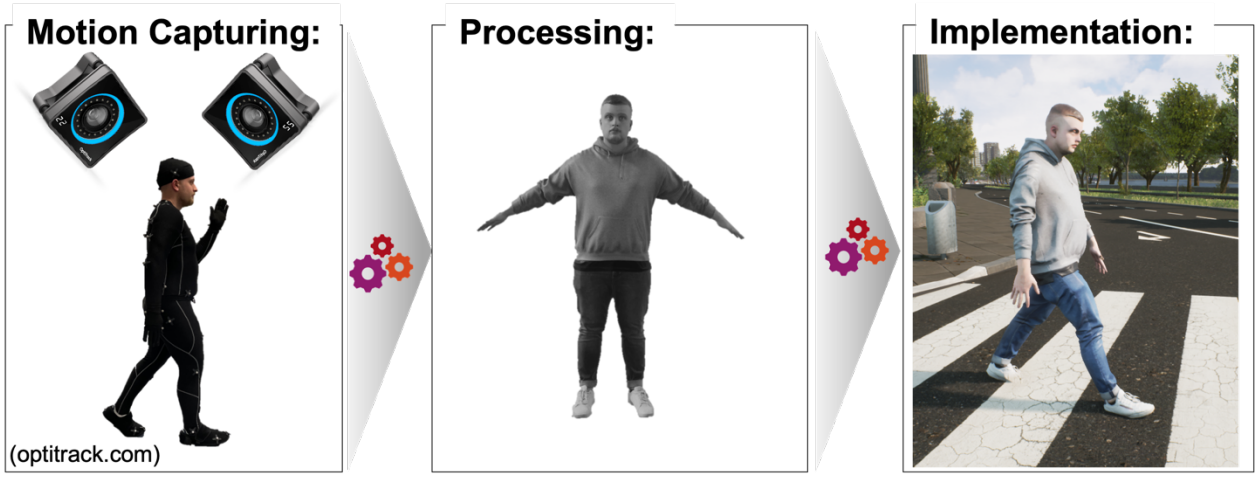

Figure 11: Processing steps for VRU-model implementation

As described before, the quality of the biomechanical capturing a representation is fine. Furthermore, the data forwarding from the motion capture hardware over the software processing to the implementation in the virtual reality (VR) environment works fluently. The movement of the VRU in the virtual scene is displayed at more than 60 frames per second (fps), which means no perceptible delay for the human eye.

5. Discussion

This paper presented two different types of motion capture based VRU modelling methodologies for the application in a virtual test field for highly automated vehicle systems. On the one hand the marker-assisted motion capturing, where PuT´s must wear a specific type of suit, which enables body movement detection in a high resolution and quality. On the other hand, the marker-less detection system, where no additional suits and markings are necessary. Both systems have individual advantages. Due to the fact, that the marker-less system needs no additional hardware, it is very fast and furthermore cheap. It uses pretrained open-source artificial intelligence algorithms to detect body joints in three-dimensional. Therefore, the accuracy is not very exact. Maybe hardware solutions with a higher number of cameras and further developed evaluation algorithms increase the detection quality and accuracy.

The marker-assisted system is expensive and the measurement technology quite complex. Thus, the distribution for the investigated use case is still quite limited. But the accuracy is very good and sufficient, and by using markers, the localization of the body joints is reproducible.

6. Conclusion

The aim of this study was to find an appropriate solution to integrate vulnerable road users' behavior into a virtual test environment for highly automated mobility systems. For this propose, the marker-assisted motion capturing methodology has turned out to be the most suitable solution. The accuracy is very good, the real-time functionality is given and the biomechanical illustration quality is fine.

Acknowledgement

The Project is supported by the Ministry of Economic Affairs, Innovation, Digitization and Energy of North Rhine-Westphalia, AVL Deutschland GmbH and by HHVISION GbR.

References

[1] **Anderhuber, Friedrich ; Pera, Franz ; Streicher, Johannes**: Waldeyer - Anato-mie des Menschen : Lehrbuch und Atlas in einem Band. 19. Aufl. s.l. : Walter de Gruyter GmbH Co.KG, 2012 (De Gruyter Studium)

[2] **Kitagawa, Midori ; Windsor, Brian**: MoCap for artists : Workflow and tech-niques for motion capture. Amsterdam : Elsevier/Focal Press, 2008

[3] **Mündermann, Lars ; Corazza, Stefano ; Andriacchi, Thomas P**.: The evolution of methods for the capture of human movement leading to markerless motion capture for biomechanical applications. In: Journal of NeuroEngineering and Re-habilitation 3 (2006), Nr. 1, S. 6

[4] **Zhang, Zhengyou:** Microsoft Kinect Sensor and Its Effect. In: IEEE Multimedia 19 (2012), Nr. 2, S. 4–10

[5] **Stewart, Kyle ; Manaris, Bill ; Kohn, Tobias:** K-Multiscope. In: COLEMAN, Grisha (Hrsg.): Proceedings of the 6th International Conference on Movement and Com-puting. New York, NY, USA : ACM, 10102019, S. 1–8

[6] **Degen, René ; Ott, Harry ; Overath, Fabian ; Klein, Florian ; Schyr, Christian ; Leijon, Mats ; Ruschitzka, Margot:** Integration of Driving Physical Properties into the Development of a Virtual Test Field for Highly Automated Vehicle Systems. NAFEMS World Congress 2021, Salzburg, 25th-29th

[7] **Degen, Rene ; Ott, Harry ; Overath, Fabian ; Klein, Florian ; Hennrich, Martin ; Leijon, Mats ; Ruschitzka, Margot:** Virtual urban traffic infrastructure for testing highly automated mobility systems. at: Krieger, J. 1. Fachkongress Digitale Transformation im Lebenszyklus der Verkehrsinfrastruktur. expert Verlag. Esslingen. 317-330. October 2021.

[8] **Müller, Meinard ; Röder, Tido ; Clausen, Michael:** Efficient content-based retrieval of motion capture data. In: ACM Transactions on Graphics 24 (2005), Nr. 3, S. 677–685