APPLYING AI METHODS ON VIDEO DOCUMENTED CAR-VRU FRONT CRASHES TO DETERMINE GENERALIZED VULNERABLE ROAD USER BEHAVIORS

Thomas, Lich
Jörg, Mönnich
Martin, Voss
BOSCH Accident Research, Corporate Sector Research and Advance Engineering, Robert Bosch GmbH,
70465 Stuttgart
Germany

Patrick, Lerge
Lennart V., Nölle
Syn, Schmitt
Institute for Modelling and Simulation of Biomechanical Systems (IMSB), University of Stuttgart,
70569 Stuttgart
Germany

Paper Number 23-0210

ABSTRACT

Urban traffic is characterized by limited traffic areas, varying traffic flows and the occurrence of different types of road users. To further advance automated mobility, the severity of injuries sustained by vulnerable road users (VRUs) in unavoidable accidents must be minimized. The project "ATTENTION", supported by the German Federal Ministry for Economic Affairs and Climate Action, was set up to tackle this issue by developing a method for the real-time prediction of VRU injury risk using artificial intelligence (AI). The present study represents the first step in the ATTENTION project and evaluates behavioral aspects of VRUs in real-life car crash scenarios. Firstly, a comprehensive, hand labeled database of video documented VRU crashes from South Korean dashcams was set up. Secondly, the data was analyzed to determine relevant characteristics like pedestrian pre-crash movement and behavior. Afterwards a comparison against the German in-depth accident study database was performed. Finally, relevant scenarios were extracted, and AI-based preprocessing was applied. Body-shape-estimation methods were used to extract pedestrian poses and kinematics for further statistical processing.

In 9,724 video documented crashes, 369 frontal primary collision against VRUs were deemed usable. The analysis reveals that every 4th crash in this sample is potentially not avoidable due to physical limitations. The VRU recognized the car before impact in every 2nd crash, possibly performing evasive actions prior to first impact. Comparisons revealed that 31,000 similar car-VRU crashes were documented in the German In-Depth Accident Study (GIDAS) database. The estimation of plausible shapes and kinematics was possible in 37 of 319 pedestrian cases (12%), while 10 of 50 videos (20%) involving cyclist could be processed. Distinct pre-crash poses and kinematics were objectively identified and were shown to be different from standard gait-cycle kinematics. The VRU shapes and poses were used to define average pre-crash body shape appearances and hull-spaces for use in future human body model simulations.

The results of this study show that a VRU pre-crash behavior can be objectively determined from low-quality in-field video data using AI-driven methods and that it differs from regular human motion patterns. Furthermore, it shows that this video data can be used to setup a position and movement database. Both lay the foundation to estimate an injury risk index of VRUs in the later stages of the ATTENTION project.

INTRODUCTION

According to the World Health Organization (WHO), about 1.35 million people die in road traffic accidents each year, with pedestrians and cyclists representing 26% of global traffic related deaths. In other words, in one out of four fatal road traffic crashes, a pedestrian or cyclists is killed [1]. Thus, ensuring the safety of vulnerable road users (VRUs) became a focus for both original equipment manufacturers (OEMs) and suppliers. Safety solutions such as the automatic emergency braking (AEB) for VRU were developed in the past and are continuously assessed in consumer ratings such as the European New Car Assessment Program (Euro NCAP) [2]. But with increasing automation, even more complex traffic scenarios involving VRUs will need to be accounted for in the design of enhanced safety systems. Modern urban traffic is characterized by limited traffic areas, varying traffic flows and the simultaneous occurrence of different types of road users. To further advance automated mobility and to address potential safety concerns related to the lack of human supervision, the severity of injuries sustained by VRUs in unavoidable accidents must be minimized. The project "ATTENTION", supported by the German Federal Ministry for Economic Affairs and Climate Action, was set up to tackle this issue by developing a method for the real-time prediction of VRU injury risk using artificial intelligence (AI) methods. The present study represents the first step in the ATTENTION project and evaluates behavioral aspects of pedestrians and cyclists in real-life car crash scenarios in terms of objective measurements like their individual shape appearance and local joint angles in the Pre-Crash-Phase.

To this end, records retrieved from event data recorders (EDRs) as well as dashcam videos collected from Korean cab drivers were first selected based on a pre-defined set of criteria and used to supplement retrospectively collected accident data, as commonly employed methods to reconstruct collisions lack information with regards to the pre-crash behavior of each crash participant. Moreover, the dashcam videos provide valuable visual information about the pre-crash phase which grants a better understanding about the behavioral aspects of the driver and other participants in the accident. The present study is the first of its kind to analyze the pre-crash behavioral aspects of Korean pedestrians and cyclists involved in frontal car collisions. To assess the applicability of the Korean traffic data thus obtained for Germany, similar selection criteria were applied to the data gathered in the German In-Depth Accident Study project (GIDAS) which provides valuable detailed information about collision events and injury mechanisms [3].

Next, the dashcam videos were further processed through AI-based body-shape-estimation methods to determine pre-collision VRU positions comparable to Schachner et al. [4] which are categorized using objective criteria based on joint angle positions. The previously published approach to derive and evaluate 3-dimensional body shapes from high definition dashcam video data is both reliant on the selection of appropriate video input and the manual classification of resulting data. The purpose of this work is to show a more automated approach that is applicable to unfiltered video data irrespective of video quality and which generates physiologically more plausible results. In addition, it is shown, how the results of the presented method can be statistically postprocessed to get a more objective general view on VRU precrash shapes, hull-spaces and kinematics.

METHODS

Data sources

The video footage used for the precrash motion reconstruction in our current study is based on data provided by the Korean Transport Institute (KOTI). In cooperation with the Korean Mutual Taxi Association (KOTMA), more than 30,000 cabs were equipped with recorders capturing video and further vehicle data [5]. The operational area of the vehicles covered mostly urban areas in Incheon City near Seoul. The video stream consists of frames with 640x480 pixels (px) resolution at a frame rate of 5 frames per second (FPS). The vehicle position was captured using the cab's Global Position System (GPS) signal and was used to derive the speed of the vehicle. Longitudinal, lateral, and vertical acceleration within a range of ± 10 m/s^2 at a 25 ms sample rate were continuously measured in addition to the video stream. In case of a collision, a minimum of 20 seconds of data-stream preceding the event were stored. Crashes of all severity levels are included, from accidents with property damage only, to crashes leading to personal injuries. No on-site investigation or any further reconstruction was done retrospectively. During an investigation period of 5 years, more than 30,000 crashes were recorded. The raw videos were provided within a collaboration of

Bosch Accident Research, KOTI and KOTMA in 2010/11. The data was provided anonymized with no personal information included. For the present study, 9,724 video documented cab crashes were available.

The study uses further in-depth accident data from the GIDAS database. Recordings contain detailed on-site information about the accident, location, and weather conditions, as well as involved parties considering more than 2,500 parameters per crash [3]. After the documentation, the cases were postprocessed and reconstructed. For the present study, we use a subset of the GIDAS database containing more than 40,000 crashes with personal injuries. To draw conclusions for Germany, GIDAS data is extrapolated using the type of crash, location, injury severity and vehicle age by applying the method described in Sulzberger et al. [6] and using data from the German Federal Statistical Office (DESTATIS) for 2021 [7][8].

Data selection
To synchronize the dashcam videos and the data from the GIDAS database, some criteria were defined and applied to both datasets. Following criteria needed to be fulfilled:

- Collision between a passenger car and pedestrian or cyclist (either 1st or 2nd participant)
- Primary impact against pedestrian or cyclist
- Impact point at the front of the vehicle (CDC2=1)[1]
- Car only moves forward (≥5 km/h, excluding standstill or backwards driving)

This allowed for the extraction of crash relevant parameters from the GIDAS database which is required to setup the parameter space for future crash simulations. After applying the listed selection criteria to the 9,724 video documented crashes, 369 frontal primary collision against VRUs were deemed usable. The selected relevant dashcam videos were processed further.

Data labeling and annotation
Firstly, a comprehensive, hand labeled database of video documented VRU crashes from South Korean dashcams was set up. Initially, the video data was not labeled or annotated, nor was further on-site investigated information about the crash available. Therefore, manual annotation of the crash sequence and extraction of relevant crash parameters was required. The capture process of various timestamps throughout the collision event is shown in Figure 1. Details about the database are available in Lich et al. [9].

Data analysis of the pre-crash behavior of VRUs
In the next step, we further analyzed the pre-crash behavior of VRUs in the remaining 369 videos based on the hand-labeled data. As the reaction of the VRUs is captured in the video image, we were able to perform a rough subjective classification. It was marked whether the VRUs perceived the vehicle before the collision or not, while the type of reaction was classified afterwards. Further details derived from the videos as well as descriptive statistics can be found in Lich et al. [9]. The VRU poses extracted from the dashcam videos will serve as time dependent target positions for the crash simulation which will follow in the later stages of the ATTENTION project, since they represent natural in crash behavior.

To describe the parameter space for the crash simulation in more detail, we analyzed the GIDAS database. To describe the pre-collision phase and in addition to commonly considered parameters such as initial- and collision speed, deceleration and impact location, the collision angle between pedestrian or cyclist and car was determined.

[1] Collision Deformation Classification (CDC) according to the National Highway Traffic Safety Administration (NHTSA)

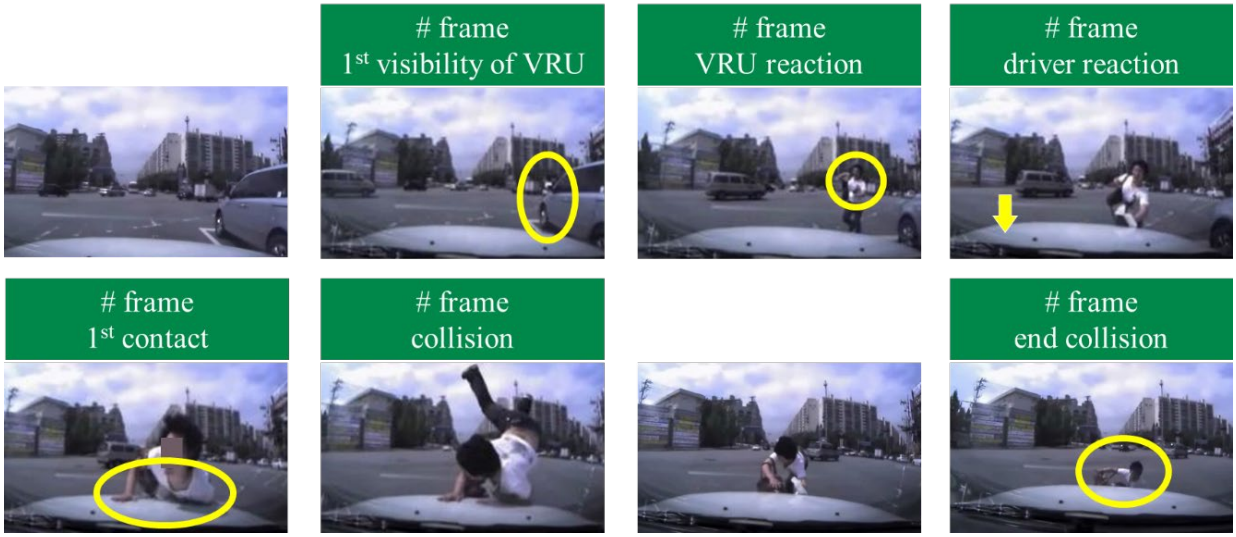


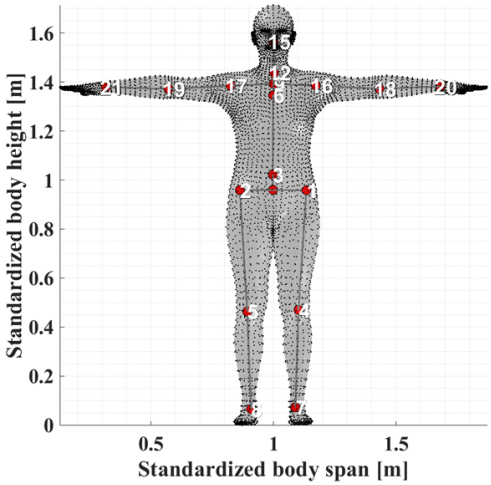
Figure 1: Capture process of various timestamps throughout the whole collision event

AI-driven Preprocessing

In parallel to the subjective analysis of the video data, an AI-driven video preprocessing was performed. First, the preselected video data was digitally enhanced to a four-times higher resolution of 2560x1920 px, with sharper contrast and less noise by applying the openly available AI-based enhancement algorithm "Real-ESRGAN" [10] with the "realesrnet-x4plus" model on all 369 videos.

These enhanced videos were processed further by applying the pose and shape estimation wrapper "PARE" [11]. PARE was chosen because it achieves a more plausible and robust estimation result when faced with partial occlusions caused by the car front or environmental structures. In this wrapper, first the 2D pose estimation software "OpenPose" [12] was applied to the videos with a tracking net resolution of a quarter the size of the video resolution. Afterwards, the resulting 2D joint locations are used as the initial condition to perform the 3D shape estimation with SMPL [13] by using the default optimization weights.

The internal skeleton regressor "SPIN", which was introduced by Kolotourus et al. [14], was used to make the underlaying skeletal structure as biofidelic and comparable to conventional motion tracking results like the H36M dataset [15] as possible. However, a modification was made to the orientation of the PARE output. The initial orientation of the internal functional skeleton was defined based on the neutral zero position instead of the original T-pose. This resulted in related movements in the joints, such as elbow and knee flexion, now being performed about the same axis. In addition, the clavicle joints and the shoulder joints have been combined into one and are only resolved in the shoulder joint. Also, the coordinate directions of the local joint coordinate systems were changed. The X-axis was made parallel to the ventral axis, the Y-axis parallel to the left-lateral axis and the Z-axis parallel to the cranial axis. All joints and their specific IDs are shown in Figure 2.



Joint ID	Joint Title	Joint ID	Joint Title
1	Left Hip Joint	12	Neck Joint
2	Right Hip Joint	15	Head Joint
3	Spine 1 Joint	16	Left Shoulder Joint
4	Left Knee Joint	17	Right Shoulder Joint
5	Right Knee Joint	18	Left Elbow Joint
6	Spine 2 Joint	19	Right Elbow Joint
7	Left Ankle Joint	20	Left Wrist Joint
8	Right Ankle Joint	21	Right Wrist Joint
9	Spine 3 Joint		

Figure 2: SMPL body shape model with SPIN joints for kinematic output in T-pose

For the results, only the shapes and poses of the crash partners were exported by manually eliminating other entities and false positive estimations that where still present after the enhancement. Apart from this, no manual correction or modification was used. The complete preprocessing workflow from the original frame to the kinematic export is exemplarily shown in Figure 3.

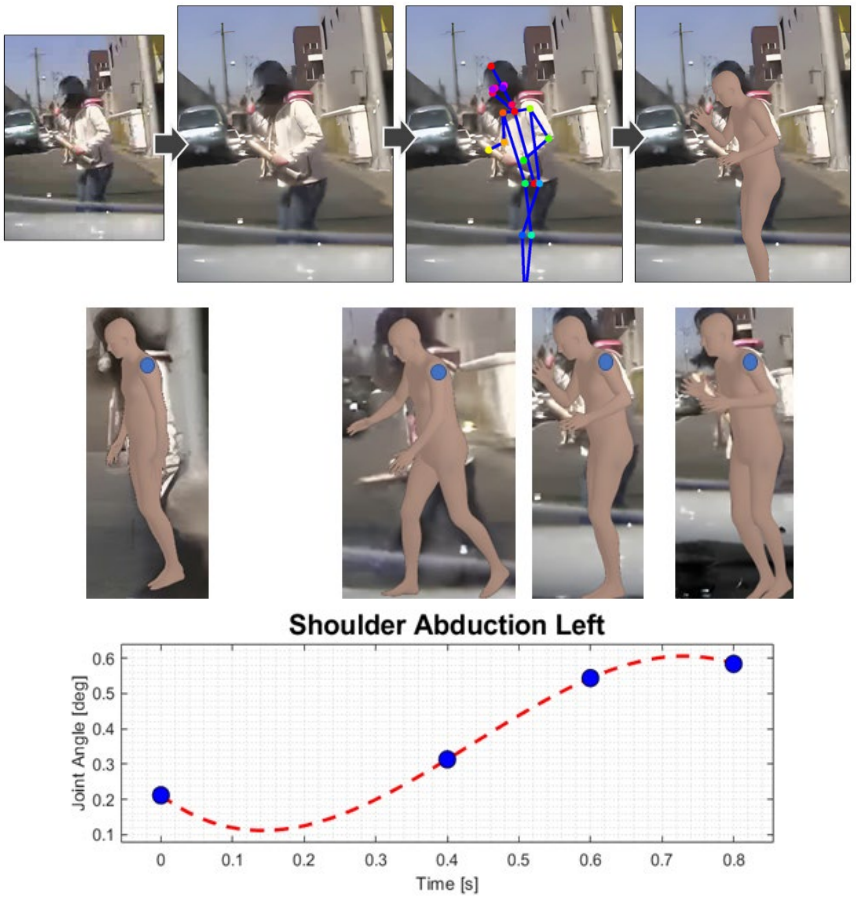


Figure 3: Preprocessing workflow: Enhancement, 2D estimation, 3D shape estimation and kinematic export

Postprocessing
In the last step, the results where statistically processed. This includes the determination of a normalized motion space over all detected entities of one type of VRU, the derivation of groups with recurring poses right before the crash, the determination of a mean shape for each group and the calculation of the average joint angles for the defined groups.

To be able to relate the appearances determined within the pose and shape estimation process to each other, a normalization of the external shapes M_N was performed. To make the outer shapes in non-local perspective comparable, all global hip-orientations (root body of the internal kinematic tree) where normalized to a 0°-orientation and all 6,890 nodes of the discretized surface M were modified with the same transformation by multiplying their XYZ-Coordinates with the 3x3 inverse matrix of the world-to-pelvis joint rotation matrix R_0 (Equation (1)), which is part of the pose estimation result.

$$M_{norm_N} = inv(R_{0,N}) * M_N \qquad \text{Equation (1)}$$

For the determination of the absolute motion space, the discretized surfaces were reduced to the normalized nodal points, and these were projected together into a three-dimensional space. Subsequently, a three-dimensional convex hull with about 400 vertices was placed around the point cloud formed in this way. For the determination of the convex hull, the method according to Barber et al. [16] was used.

Thus, a theoretical motion space could be identified, which, applied backwards to each individual orientation of an entity in relation to the world, provides information about a possible collision volume around a person for which only the orientation and movement of one point (in this case of the hip) must be predicted.

In the next step, the behavioral groups were determined. For this, the local joint angle differences from the final frame right before the impact were calculated first. To do so, each of the i = 23 joints of one entity 'm' was compared to its equivalent of another entity 'n' by determining the absolute angular difference for each of the 3 degrees of freedom (DOF) joints ($\theta_{diff\,i,mn}$) using the 3x3 rotation matrices R_i that represent these local joint angles as shown in Equation (2).

$$\theta_{diff\,i,mn} = \cos^{-1}\left(\frac{tr(R_{i,m}R_{i,n}^T)-1}{2}\right) \qquad \text{Equation (2)}$$

This lead to a difference-vector with i = 23 elements for every entity-to-entity relation, which was reduced to a mean angular error using the Root-Mean-Square method in the next step. This, for N entities, resulted in an error-matrix of the dimension NxN and was finally normalized by dividing the mean error values by the maximum possible error of 180°, which created the normalized error-matrix θ_{error} (Equation (3)).

$$\theta_{error_{mn}} = \frac{1}{180°}\sqrt{\frac{1}{k}\sum_{i=1}^{k=23}\theta_{diff_{i,mn}}^2} \qquad \text{Equation (3)}$$

Finally, network analysis was used to create a network comprised of those entities, whose error was non-zero and less than a variable generalized threshold τ. To find appropriate values for τ , it is recommended to perform an optimization to identify the value at which the number of groups of entities with more than one member is at a maximum and the number of entities without connection is at a minimum.

In the next step and based on the groups found this way, an average shape regarding every entity in one group was determined by calculating the average 3D-Position of every surface node over all entities in this group 'G'. The number of existing connections from the network 'deg' was also included as a weighting factor (Equation (4)).

$$M_G = \frac{1}{\sum_{i_G}^{N_G} deg_{i_G}}\sum_{i_G}^{N_G}(deg_{i_G}\, M_{norm_{i_G}}) \qquad \text{Equation (4)}$$

In this way, the number of G discretized shapes could be determined for the different final appearances in the pre-crash phase. However, with the data available in this study, not only the generalized outer shapes but also group-specific average joint angles could be determined by using the output of the functional skeleton, which is a function

of the body shape. For this purpose, the i = 23 joint rotation matrices of each entity in a group N_G were converted into quaternion form. This led to N_Gx23x4 sized matrices. Then, the number of occurrences of each entity in the matrix was multiplied according to its number of connections in the group-network/ level of 'deg' to get the same weighting effect as it was taken into account for the average shape.

These extended quaternion matrices were further averaged according to Markley et al. [17] which finally resulted in a matrix of size 23x4 for each individual group. At this point, the approach via quaternions had to be taken, since the supposedly intuitive cardan angles are not explicit due to sequencing in XYZ rotations, but are instead corrupted by previous rotations in their sequence and can therefore not be used for elementwise calculations. This matrix was then transformed back into rotational matrix formulation of size 23x3x3, as well as into a sequential X-Y'-Z'' cardan rotation.

The post-processing methods described above were applied to both the reconstructed videos of pedestrians and cyclists. However, both groups were always considered as separate types of VRUs and as such processed independently.

RESULTS

Field of effect

Applying the previously outlined selection criteria to the GIDAS database and extrapolating it towards the entirety of German traffic reveals that the selected Korean traffic scenarios correspond to about 29,000 similar car-VRU crashes with personal injuries in Germany. Overall, this is a share of about 11% of all car-VRU front crashes with personal injuries in Germany in 2021. Table 1 shows the selection process for both data sources.

Table 1: Available data and relevance for Germany

Criteria	Number of video-documented crashes (Korea 2010/11)	Estimated represented number of crashes in Germany (2021)
All crashes w/ personal injuries	9,724	258,987
... and car involvement	9,563	192,221
... and pedestrian or cyclist involvement	641	~ 66,000
... and first collision w/ cyclist or pedestrian	641	~ 60,000
... and car, pedestrian, cyclist as 1./2. participant		~ 59,000
… and frontal car collision (CDC2=1)[2]	375	~ 30,000
… and car collision speed min. 5 km/h incl. unknown (Field of effect of pfp ATTENTION)	369	~ 29,000
… out of pedestrian crashes	319	~ 8,000
… out of bicycle crashes	50	~ 21,000

Result Pre-crash behavior for pedestrians from video-data

For pedestrians we observed that in about 43% of the 319 sample the vehicle was not recognized prior to the collision whereas in 48% the pedestrian saw the car as shown in Figure 4. Once the pedestrian perceived the car, some reaction takes places. The decision-making process was not further evaluated. Moreover, the aim was to determine the parameter space which is needed to cover pedestrian crashes on a wider scope. However, in 38% of these cases, the pedestrian recognizes the car but did not change his behavior. Another 31% stopped or slowed down, while 16% sped up. In 10%, other movements took place, and in 3%, a defense posture was observed. These findings were further confirmed once the poses were extracted out of the video images.

[2] Collision Deformation Classification (CDC) according to the National Highway Traffic Safety Administration (NHTSA)

In Lich et al. [9], the time to collision (TTC) was determined. As a result, it was found that in about 25% of the accidents the time to collision is less than 1.2 seconds. This means that in one out of four car-pedestrian accidents, the collision cannot be prevented in this observed sample.

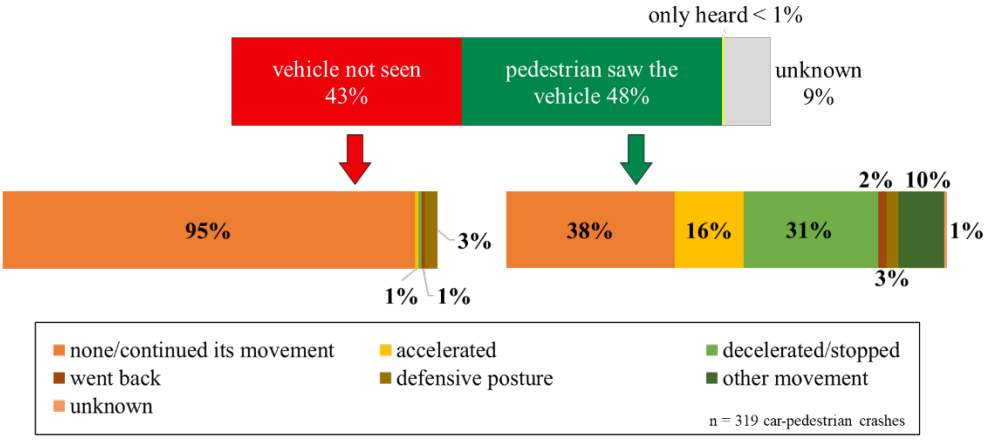


Figure 4: Pedestrian reaction by recognition prior to the impact determined out of 319 video documented car-pedestrian crashes

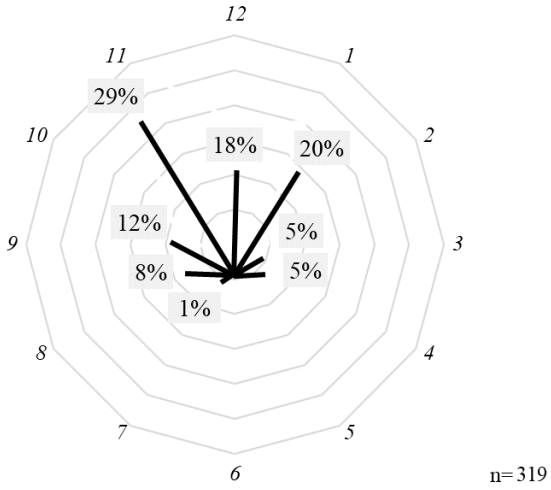


Figure 5 : Pedestrian impact direction determined out of 319 video documented car-pedestrian crashes

The impact directions of the pedestrians against the cars are shown in Figure 5. Most of the front crashes occur at impact positions of 11 o'clock (29%), 1 o'clock (20%) and full frontal at 12 o'clock (18%).

Result Pre-crash behavior for cyclists from video-data

In the sample of 50 car-cyclist crashes, 42% of cyclists did not see the car coming as in some of the crashes, the cyclists were hit from behind and thus were not able to recognize the car at all. However, 52% of them perceived the car but showed no reaction (77%). Some of them accelerated (8%) to avoid the collision and another 8% evaded mainly by steering intervention. The remaining 8% performed a mixed motion of steering and evading. Figure 6 shows the findings for the cyclist's pre-crash behavior.

Furthermore, for car-bicycle accidents, it was found that in about 18%, the TTC is less than 1.2 seconds. This means that in one out of five car-bicycle accidents, the collision cannot be prevented in this observed sample. The TTC was evaluated in detail in Lich et al. [9].

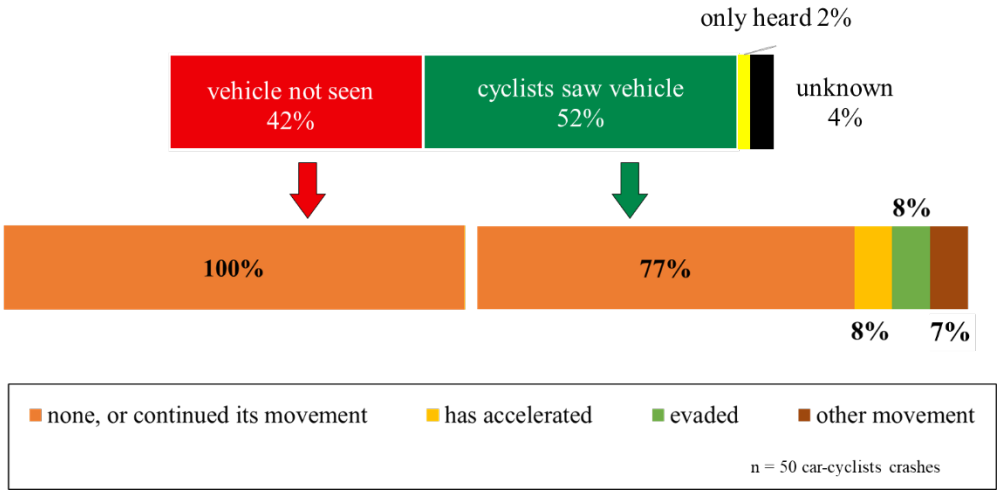


Figure 6: Cyclist reaction by recognition prior to the impact determined out of 50 video documented car-cyclist crashes

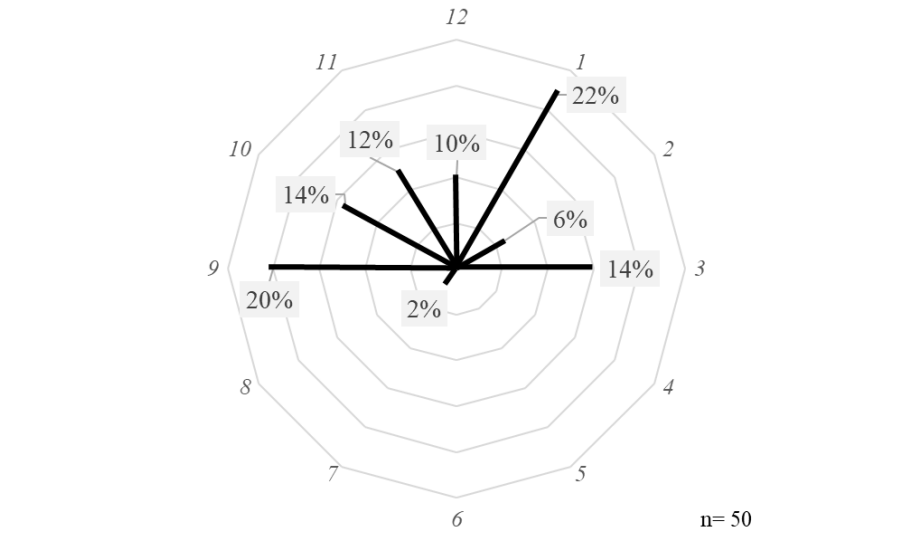


Figure 7: Cyclist impact direction determined out of 50 video documented car-cyclist crashes

The impact directions of the cyclists against the cars are shown in Figure 7. The majority of crashes occur at 1 o'clock (22%). It can also be seen that in 1 out of 3 crashes, the cyclists crossed the path of the vehicle laterally.

Results Enhancement

The initial use of an enhancement algorithm led to an improvement in pose and shape estimation such that the initial two-dimensional detection of people using OpenPose started on average 2 frames earlier (0.4 s; ~50% of total length) as shown in Figure 8. At the same time, the number of false positive detections decreased by ~20% due to the AI-assisted resharpening so that longer and qualitatively more reliable kinematics were achieved.

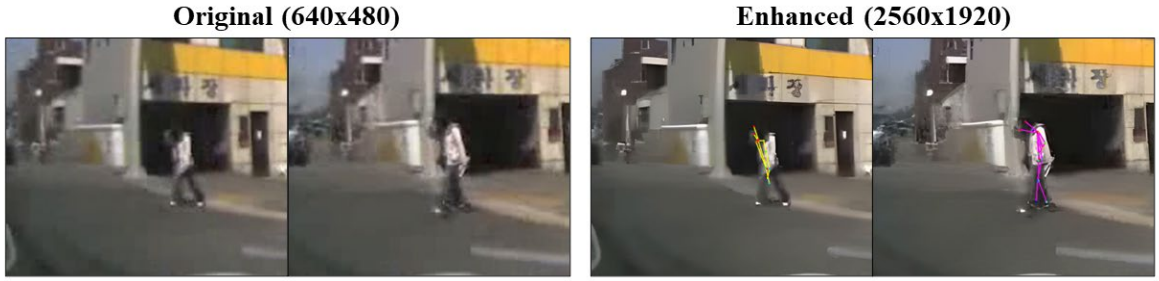


Figure 8: Effect of video enhancement on 2D pretracking with OpenPose

Results Person specific precrash shapes and poses

Starting from the database of preselected videos that represent our selected load case, an acceptable estimation of plausible shapes and kinematics was possible in 37 of 319 pedestrian cases (12%), while 10 of 50 videos (20%) involving cyclist could be processed. As illustrated in Figure 9, the behavior of natural individuals right before a crash with a vehicle, as they were estimated with the presented toolchain, is highly diverse. Most of the randomly selected individuals here show a behavior that is different from natural movements beside accidents like a natural locomotion. It is therefore highly advisable to define objective conditions that can be used to define and classify the different types of behavior in a more simplified yet appropriate form.

Figure 9: Examples of non-classified pedestrian postures prior to the collision derived from the video documented crashes

Results Hull-Space Pedestrian

Since the individual appearances, as shown in Figure 9, vary widely and appear to be unequally directed, the shape orientation was normalized in the first step of statistical processing. By applying the same origin point to all entities, a situation dependent maximum extension of the human body can be determined, which is unequal to the general maximum possible expansion of the human body, without the need to consider local joint angles in the first place. The resulting motion/hull-space as a result of all our reconstructed shapes is displayed in Figure 10. The hull allows for a rough estimation up until which point a collision is possible and where it is unlikely, only considering the hip centre position, size scaling and hip orientation.

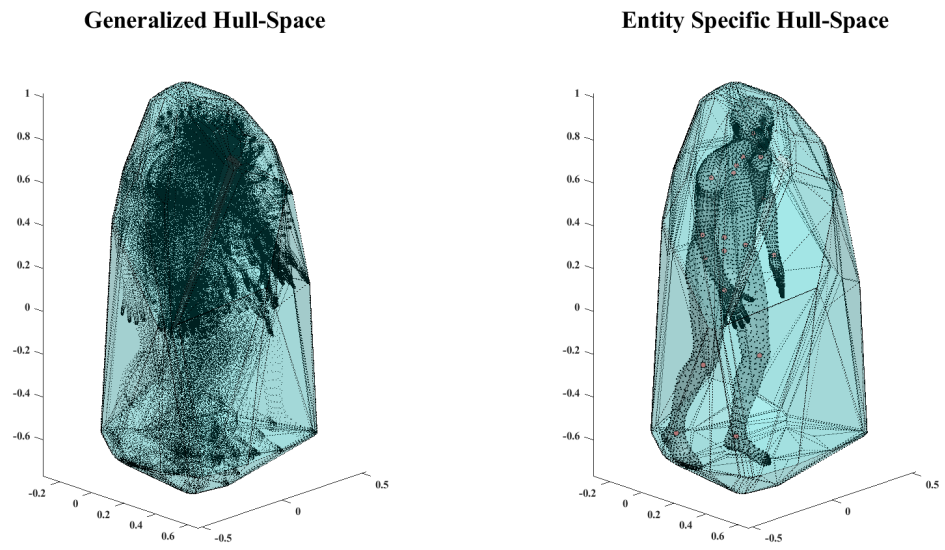


Figure 10: Observed hull-space over all pedestrians and for one representative example shape

Results Grouping Pedestrian

Due to the still very different moving appearances, a method was applied to group the reconstructed shapes and the associated movements and to generalize the external appearance as well as the movement at the level of joint angles in a group-specific way. These mean shapes and poses should still be diversified enough to be used for the further definition of pre-crash target positions for simulations with human body models.

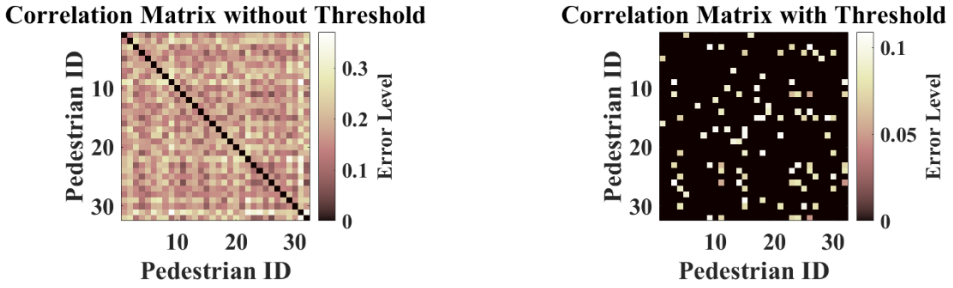


Figure 11: Pedestrian correlation matrix

The matrices shown in Figure 11 are quasi-correlation matrices there a larger cell value represents a larger mean difference in the joint angles and thus a stronger deviation in the outer appearance. The general purpose of these matrices is to find entity relationships with a high level of similarity, which was implemented by applying a threshold that annulled all cells with a larger error relation level. In the present case, a uniform threshold of 10.9% was used, as this fulfilled the criterion for a maximum number of groups with more than one entity at a maximum group size. The mentioned groups are shown in the following Figure 12.

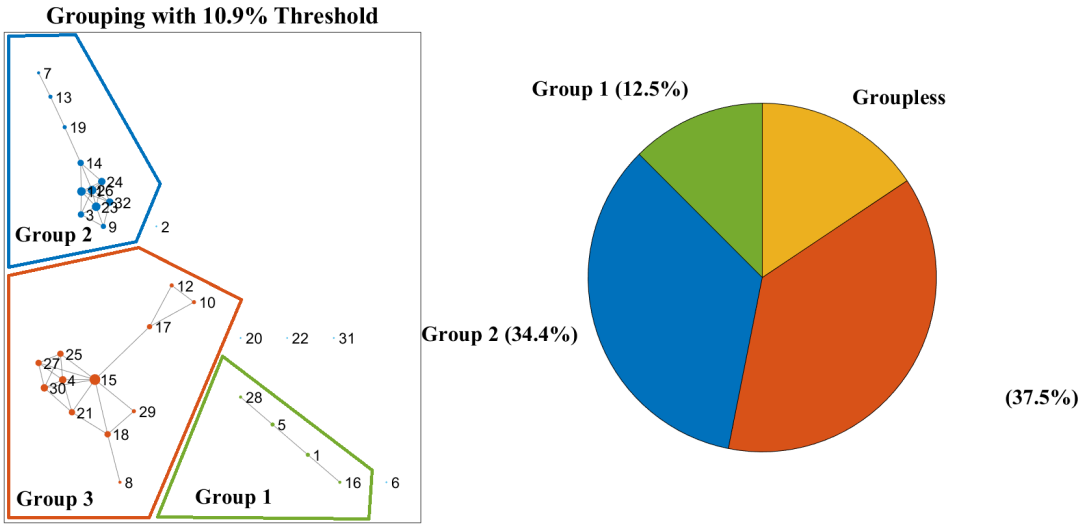


Figure 12: Resulting groups of pedestrian behavior

Figure 12 visualizes the distribution of the pedestrian entities in relation to each other as a graph and as a pie chart. With the selected threshold, three larger groups can be identified that are self-contained and thus visually distinguish themselves from the other appearances. Within these groups, there is at least one entity that has a particularly high similarity to most of the other group members, which should be taken into account in the further averaging of the body shapes.

It can be observed that the formed groups are not uniformly sized. Accordingly, there is a more frequent and a less frequent behavior in a crash situation. For example, group 3 comprises 37.5% of all pedestrian cases, while group 1 comprises only 12.5%. In total, 84.4% of the reconstructed cases were captured via the grouping method, which were subsequently included in the following shape and pose normalization.

Results Shape and Pose Pedestrian

In the following Figure 13, only the mean outer shapes and the joint locations are shown. The influence of each group entity was weighted depending on the number of its connections in the network.

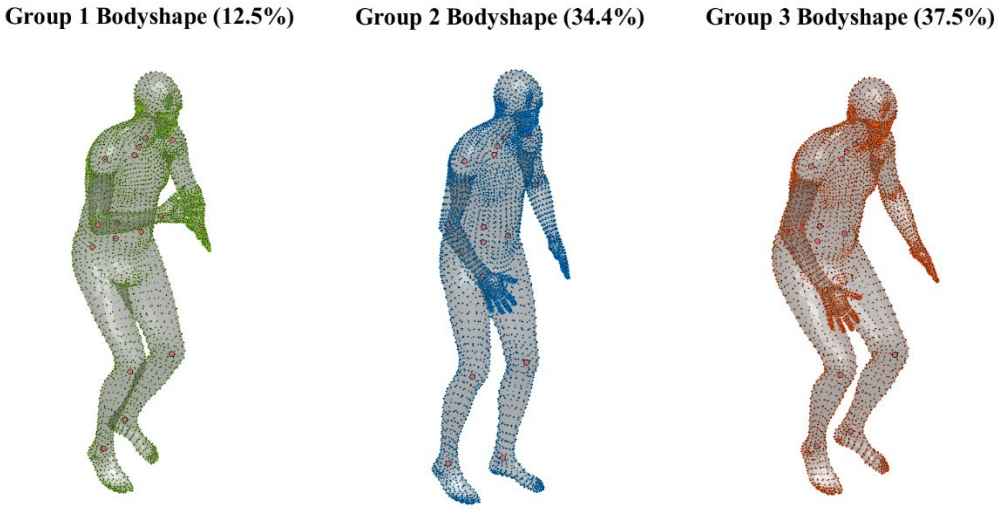


Figure 13: Mean pedestrian pre-crash shapes for groups larger one subject

Compared to Figure 9, the illustrations in Figure 13 show a much more symmetrical and less distorted appearance. The smallest group, consisting of 4 people (12.5%), is characterized by the fact that the respective entities have adopted a protective or defensive posture with their arms up and an angled leg position.

The average shape of the members of the largest Group 3 however, shows a body posture that corresponds more to a fright, preparation, or flight posture. It is characterized by angled legs like in group 1 but with a smaller inside angle as well as downward pointing and opened arms.

In contrast, the posture of group 2 barely deviates from a natural body posture. It can therefore be assumed that in at least 34.4% of our cases, there was no or barely any preparatory reaction to the impending accident.

Table 2: Pedestrian mean pose joint angles

	Group 1			Group 2			Group 3		
JOINT ID	**X [°]**	**Y' [°]**	**Z'' [°]**	**X [°]**	**Y' [°]**	**Z'' [°]**	**X [°]**	**Y' [°]**	**Z'' [°]**
1 (L Hip)	5.84	-24.14	0.68	1.40	-17.79	0.21	8.52	-37.77	2.56
2 (R Hip)	-4.16	-30.94	-1.94	-3.58	-10.81	-1.09	-6.94	-39.52	-5.01
3 (Spine 1)	0.24	20.03	-0.96	0.70	14.75	0.91	0.56	25.31	-0.31
4 (L Knee)	-7.87	54.19	2.03	-4.27	24.39	0.50	-14.05	61.94	9.74
5 (R Knee)	9.18	51.38	-5.45	2.22	24.90	1.45	20.97	62.55	-18.83
6 (Spine 2)	0.22	1.41	-0.01	0.27	2.08	0.33	0.34	1.19	0.46
7 (L Ankle)	-6.44	-1.31	10.52	-6.22	0.49	8.43	-5.26	-0.58	10.14
8 (R Ankle)	6.42	1.02	-8.33	6.53	1.79	-7.79	5.39	4.01	-8.21
9 (Spine 3)	-0.69	2.72	-0.76	-0.37	3.12	0.59	-0.64	3.23	0.02
12 (Neck)	0.71	-3.21	0.14	-0.24	-0.11	-1.41	-0.30	-1.69	-3.97
15 (Head)	-0.52	4.71	2.61	-1.23	3.23	0.57	0.96	2.15	-3.35
16 (L Shoulder)	17.65	-34.67	-39.10	19.93	-11.43	-22.13	21.29	-26.18	-28.92
17 (R Shoulder)	-19.06	-35.67	35.67	-16.40	-16.90	26.93	-18.22	-25.08	30.78
18 (L Elbow)	43.10	-75.92	24.56	0.19	-49.20	-12.44	21.29	-67.69	4.93
19 (R Elbow)	55.88	-110.06	75.13	3.01	-47.71	14.55	-10.22	47.39	2.03
20 (L Wrist)	10.30	-3.99	2.71	8.44	-4.67	2.27	11.44	-3.87	0.02
21 (R Wrist)	-11.63	-5.19	-3.37	-9.55	-4.29	-2.03	-10.69	-3.21	1.01

The scaling invariant joint angle description of the three poses is listed in Table 2. Again, a weighting was applied to the determined pose data. These calculated joint angles are listed in the cardan notation for easier interpretation.

The values listed in Table 2 allow for an additional check for plausibility by comparing the angles at the supposedly natural joints in the hip, knee, ankle, shoulder, and elbow with limit values from the literature such as those in Barter et al. [18]. In this case, all joint angles passed this test. The reconstructed shapes and poses determined from them are therefore physiologically plausible.

Results Hull-Space Cyclist

After the highly diverse pedestrians, the hull and grouping methodology was also applied to the reconstructed data of cyclists. However, it neglects the frame of the bicycle which is not estimated during the pose and shape estimation but still a contactable object and effectively connected to the rider.

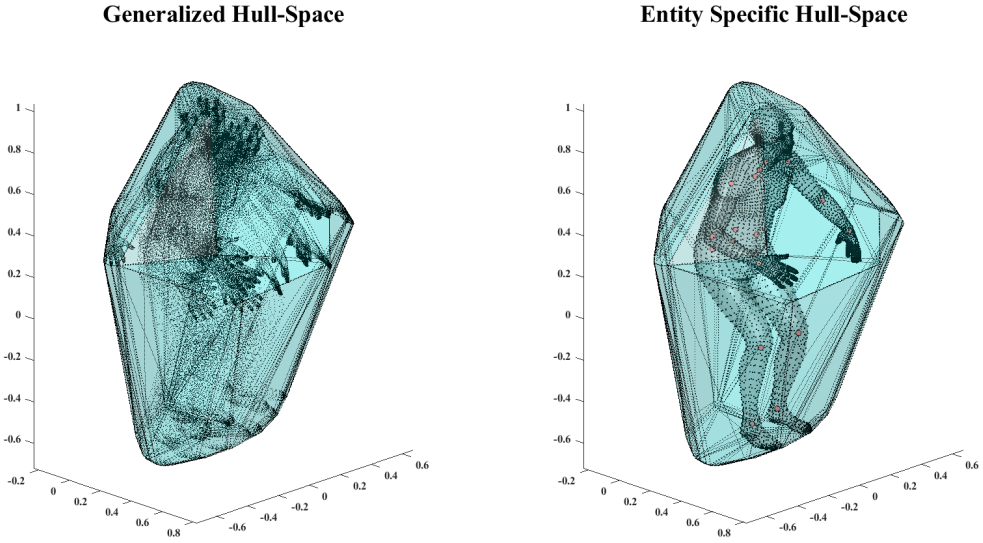


Figure 14: Observed motion space over all cyclists and for one representative example shape

The hull in Figure 14 strongly indicates the influences of the forced seating position with the two feet firmly on or at the pedals. At the same time, it can be read from the standardized width scale that a cyclist needs more space in the width dimension than the pedestrian, which is not only due to the nature of the handlebars, but also to perform driving related gestures like hand signals.

Results Grouping Cyclists

The left quasi-correlation matrix from Figure 15 already shows a nearly 50% lower maximum error level compared to the distribution of pedestrians from Figure 11.

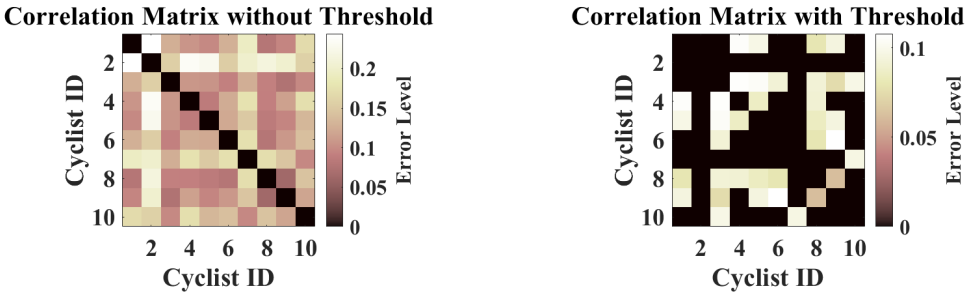


Figure 15: Cyclist correlation matrix

The results from Figure 16 support the initial interpretation of the lower error level from the previous figure. There was no threshold at which more than one group with more than one group member would have formed. Therefore, for comparability, the same threshold was chosen as for the pedestrian. Along with this, one large group was formed, with only a single reconstruction not adhering to the group resemblance.

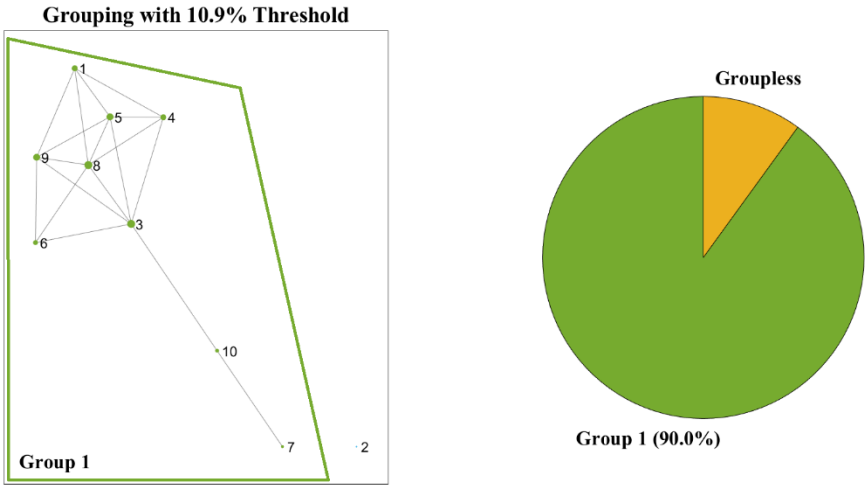


Figure 16: Resulting groups of cyclist behavior

Results Shape and Pose Cyclist

The mean shape from the grouped cyclist data is shown in Figure 17.

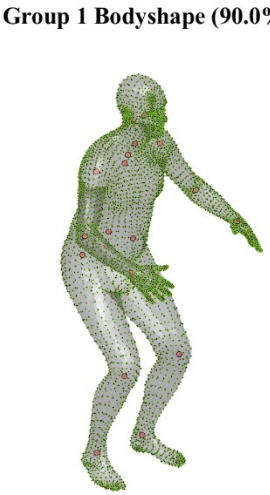


Figure 17: Mean cyclist pre-crash shape for group larger one subject

The shape subjectively matches that of a cyclist, but it projects legs and feet parallel to each other. This may have been caused by the applied averaging methods or by an error in the computer-vision-based shape reconstruction.

In general, there is no evidence of any special behavior of cyclists that differs from the natural driving behavior.

The joint angles of the singular mean cyclist position, as they are listed in Table 3, do not show any irregularities at all. Similar to the pedestrians, the cyclists' pose also withstands the test for biomechanical physiological plausibility. As such, they form a plausible and objectively tangible foundation for future model driven test case studies with cyclist VRUs.

Table 3: Cyclist mean pose joint angles

	Group 1		
JOINT ID	**X [°]**	**Y' [°]**	**Z'' [°]**
1 (L Hip)	9.79	-36.07	0.42
2 (R Hip)	-8.79	-37.82	-3.74
3 (Spine 1)	0.48	24.49	1.25
4 (L Knee)	-14.95	55.10	3.68
5 (R Knee)	14.63	58.41	-11.23
6 (Spine 2)	-0.06	-3.61	-1.05
7 (L Ankle)	2.92	-14.04	8.20
8 (R Ankle)	4.45	-13.93	-10.33
9 (Spine 3)	1.60	-1.29	1.62
12 (Neck)	-4.12	-5.04	6.12
15 (Head)	4.56	-3.47	2.24
16 (L Shoulder)	37.53	-34.33	-15.37
17 (R Shoulder)	-39.35	-34.09	17.28
18 (L Elbow)	21.49	-58.02	4.28
19 (R Elbow)	-19.40	-54.28	-7.64
20 (L Wrist)	-0.91	-8.58	0.97
21 (R Wrist)	-0.74	-8.86	-5.53

DISCUSSION

The work presented in this submission is subject to several limitations. The GIDAS database gathers in-depth information about accidents with casualties weighted for Germany, which might not strictly apply to the video database from South Korea. Moreover, the video data shows that such information is rather essential when it comes to the behavioral aspects of VRU during the pre-crash phase, as such information cannot be retrospectively gathered from on-site investigations. This in-depth data would help to further specify the mechanisms of injuries sustained and to reconstruct the vehicle speed. All in all, both data sources would allow for a wide range of accident crash causation analysis.

The video footage used in this study is characterized by a low sample rate and a low resolution. Some misinterpretation of the behavioral aspects might therefore occur during the kinematics estimation processes. Due to the low sample rate, components of a movement may also have been lost because the movement signal was under sampled. Additionally, there is no "ground truth" data available which could be used to assess the validity of the estimations. This point was addressed at least to the extent that the mean postures were tested for physiological plausibility at the joint angle level.

An additional uncertainty is introduced through the missing representation of the individual body segment lengths of every subject. The used shape model scales its volume but not the average length of each body segment. This leads to a constant length relation of upper- to lower leg or torso length to leg length. This is not the case in nature and might result in an additional error regarding the estimated joint location. Also, the used shape model is not scalable in its total height. This affects mainly the volume of the estimated hull which presently only represents a 50-percentile male. It would thus be necessary to scale the hull volume relative to the real body height of the pedestrian or cyclist involved in the accident. The additionally collected kinematic data is not affected by this issue since they are in general scaling invariant.

Nevertheless, the general shape estimation from frame data is still the best method to obtain plausible information from real live video material. As the quality of the underlying data may increases, so will the quality of the reconstruction in future applications. Applying the method to high resolution data may however not require

additional preprocessing to improve the image quality. Still, the study shows that even with a low sampled video stream, a pose and motion database can be setup without necessitating high end video hardware.

The sample rate played only a minor role in the statistical grouping applied in this work. The authors have successfully found a method to classify different types of VRUs in a time-efficient and objective way without having to train additional neural networks in advance, which makes the method very transparent as well. Nevertheless, this method also offers potential for optimization and further development. One obvious weakness is that the evaluated pose similarity is based on only one frame. With a better sampling, however, the correlation of the kinematics over time would be much more informative. In addition, it is debatable whether all joints should be equally important for the similarity evaluation or whether they should be weighted according to a criterion such as the distance to the center of the body. Finally, this work shows that average poses for VRUs could be determined but their influence on the outcome of a crash is yet unclear. For the final differentiation of tests with post-mortem human subjects and passive models, it would now be necessary to determine what distinguishes the poses on a dynamic level and how they differ from each other on the force and momentum side. However, this could be addressed with muscle-driven human body models and the data we have identified as a target variable in future studies.

Applying the ATTENTION criteria to Germany, the study furthermore reveals that the priority should be set on car crashes against cyclists as these are represented with a share of 8% of all crashes with personal injuries.

As the sample shows, in nearly every 2nd crash (43%) the VRU did not recognize the oncoming vehicle. However, in case of perceiving the vehicle, different reaction patterns were observed. These patterns can be further evaluated using objective criteria such that a classification of the patterns is made possible. A similar behavioral pattern is also found in the relative group distribution across the pose grouping. Here, little to no visible reaction to the upcoming crash was found for at least 34.4%. Thus, the subjective and objective analyses come to a comparable conclusion in terms of their magnitude. In addition, 15.6% of all cases still were without inferred motion intention. It is not yet possible to say whether these are generally unusual behaviors, if their outer appearance was strongly influenced by an additional attraction or whether this is a phenomenon based on the small sample size.

With the cyclists on the other hand, the distribution of physical appearance was much less diverse, resulting in a single group instead of three. At the same time, the subjective evaluation of the behavior was much more difficult than with the pedestrians, which was also due to the reduced group size. Considering the result that there is no special precrash defense motion for cyclists, we conclude that the decisive influence for the outer shape of a cyclist is the type of bicycle which is being ridden. Thus, the only person who fell out of the presented grouping was riding a road bike where otherwise roadsters were present in the videos. In this context, the result represents a good and objective reference for those who are interested in the reproduction of natural movements and postures on a bicycle. Nevertheless, it can also be seen that cooperative crash avoidance by the VRUs cannot be assumed when it comes to the development of crash avoidance systems.

For both samples the determination of the impact angle was also rather limited as no objective criteria were applied on the video data. Nonetheless, the results give some valuable input when it comes to the impact itself as a crossing VRU will have different injury patterns compared to impacts in which the VRU is hit head-on.

The objective analysis methods presented in this work can help to improve behavior categorization. Generalized and scaling invariant joint angles can be further used for simulative situation analysis via pose and motion mimicking in a situation reconstruction. A typical approach to measure such kinematics is via marker-based motion capturing, which is not possible (for ethical reasons) for the relevant use case. We are therefore limited to marker-less tracing methods, that are applicable retrospectively to make the most of past recordings such as the available dashcam videos. Since these are only available as 2D data, but 3D information is mandatory to reconstruct the whole body's position and local joint angles, we are limited to methods that are using dimensional estimation strategies via AI to bypass the redundancy problem. One possible approach that was already used by Technical University of Graz (TU Graz) [4] for their behavioral categorization, is the application of a multidimensionally scalable human shape model which allows to reconstruct not just 3 translational but 3 rotational dimensions in the current use case as well.

CONCLUSIONS

The results of this study show that a pre-crash behavior of VRUs can be objectively determined from low-quality in-field video data using AI-driven methods and that it differs from regular human locomotion patterns. Furthermore, it shows that this video data can be used to set up a position and movement database. Both lay the foundation to estimate an injury risk index of VRUs in the following stages of the ATTENTION project.

State-of-the-art accident research uses data from post-accident investigations which is gathered in databases like GIDAS. Information about the impact moment may be inferred but the pre-crash behavior of VRUs is usually out of reach for these kinds of investigations. To the authors' knowledge, this paper presents the first systematic reconstruction of how people behave before the crash and react to the imminent impact, e.g., do they react at all or change their behaviors in distinct ways. This data can be used to supplement retrospective data which may lead to a better understanding of crash causes, injury patterns and new safety solutions from the automotive industry for protecting VRUs. Therefore, the dashcam video data is a valuable additional asset. Currently, the amount of video data analyzed is limited. To complete the picture of the behavioral patterns of VRUs prior to impact, more data should be added to form a comprehensive database.

Finally, and thanks to its automated and generalized nature, our method should be applied to a larger amount of dashcam video data to increase the statistical significance of this work. The great advantage of our method is that it is possible to objectively evaluate and compare large amounts of results in a short time.

Within the project ATTENTION, the behavioral pre-crash patterns of VRUs will be used as an input for crash simulations. The severity of the injuries sustained by the VRUs within these crash simulations will be determined and mapped to the behavioral patterns which will form a new database for crash analysis. The final goal within the project ATTENTION is to deduce the potential injury risk of a VRU before an accident, based on the detection of these, their behavior before the impact as well as the trajectory of the involved vehicle. This may give rise to new driving functions using an optimized combination of evasive steering and emergency braking to protect the VRU in an unavoidable accident as well as possible. This would provide further insight into which avoidance or mitigation strategies can be initiated in case of unavoidable VRU accidents.

ACKNOWLEDGEMENTS

The project "ATTENTION" (long title: Artificial Intelligence for real-time injury prediction) on which this report is based was funded by the German Federal Ministry of Economic Affairs and Climate Action (Bundesministerium für Wirtschaft und Klimaschutz) under the funding code 19A21027A.

REFERENCES

[1] Global status report on road safety 2018. Geneva: World Health Organization; 2018. Licence: CC BYNC-SA 3.0 IGO. ISBN 978-92-4-156568-4

[2] European New Car Assessment Programme (EuroNCAP) – Test Protocol AEB/LSS VRU systems, Implementation 2023, Version 4.3, November 2022, accessed on November 28th 2022 at www.euroncap.com/en/for-engineers/protocols/vulnerable-road-user-vru-protection/

[3] Liers, H.. 2018. "Traffic accident research in Germany and the German in-depth accident study (GIDAS)", SIAM Conference

[4] Schachner, M. and Schneider, B., and Klug, C. and Sinz, W.. 2020/9/1. "Extracting quantitative descriptions of pedestrian pre-crash postures from real-world accident videos", International Research Council on Biomechanics of Injury (IRCOBI).

[5] Sul, J. and Cho, S.. 2009. "Obtaining and applying of traffic accident data using automatic accident recording system in Korea". 4th International Road Traffic Accident Database (IRTAD) conference, Seoul. Korea. Pp.-395-395

[6] Sulzberger, L. and Schmidt, D. and Moennich, J. and Schlender, T. and Lich, T. 2022. “How to use historic accident data for a reliable assessment of traffic safety measurements”, 7th International Road Traffic Accident Database (IRTAD) conference, Lyon, France.

[7] Federal Statistical Office of Germany (editor). August 2021. Special evaluation of traffic accident 2020 on behalf of Bosch Corporation

[8] Federal Statistical Office of Germany (editor). 2021. “Verkehrsunfälle 2020, Fachserie 8, Reihe 7”. www.destatis.de

[9] Lich, T. and Moennich, J. and Schmidt, D. and Voss, M.. 2022. “Preparation of an AI based real-time injury risk index estimation by deriving road user behavior from video-documented crashes”, 15th International Symposium and Exhibition on Sophisticated Car Safety System (airbag2022), Mannheim. Germany. Fraunhofer ICT (editor). ISSN 0722-4087. Pp. 20.1-20.19.

[10] Wang, X. and Yie, L. and Dong, C. and Shan, Y. 2021. “Real-ESRGAN: Training Real-World Blind Super-Resolution With Pure Synthetic Data” In Proceedings of the IEEE/CVF International Conference on Computer Vision, 1905-1914

[11] Kocabas, M. and Huang, C.-H. and Hilliges, O. and Black, M. 2021. “PARE: Part Attention Regressor for 3D Human Body Estimation” In Proceedings of the IEEE/CVF International Conference on Computer Vision, 11127-11137

[12] Cao, Z. and Simon, T. and Wei, S.-E. and Sheikh, Y. 2017. “Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields” In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 7291-7299

[13] Loper, M. and Mahmood, N. and Romero, J. and Pons-Moll, G. and Black, M. 2015. “SMPL: a skinned multi-person linear model” ACM Transactions on Graphics 34, No. 6, 248:1-248:16

[14] Kolotouros, N. and Pavlakos, G. and Black, M. and Daniilidis, K. 2019. “Learning to Reconstruct 3D Human Pose and Shape via Model-Fitting in the Loop” In Proceedings of the IEEE/CVF International Conference on Computer Vision, 2252-2261

[15] Ionescu, C. and Papava, D. and Olaru, V. and Sminchisescu, C. 2014. “Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, No. 7

[16] Barber, C.B. and Dobkin, D.P. and Huhdanpaa, H. 1996. “The quickhull algorithm for convex hulls” ACM Transactions on Mathematical Software 22, No. 4, 469-483

[17] Markley, F.L. and Cheng, Y. and Crassidis, J.L. and Oshman, Y. 2007. “Averaging Quaternions” Journal of Guidance, Control, and Dynamics 30, No. 4, 1193-1197

[18] Barter, T. and Emmanuel, I. and Truett, B. 1957. “A statistical evaluation of joint range data” WADC Technical Note 53-311, Wright Patterson Airforce Base, OH