BRAVE

BRidging gaps for the adoption of Automated VEhicles

No 723021

D4.1 Methodology for Vehicles and VRUs prediction of intentions

Lead Author: Miguel Ángel Sotelo (UAH)

With contributions from: Ignacio Parra, David F. Llorca, Carlota Salinas, Javier Alonso (UAH)

Deliverable nature:	Report (R)
Dissemination level: (Confidentiality)	Public (PU)
Contractual delivery date:	28-February-2018
Actual delivery date:	12-February-2018
Version:	1.0
Total number of pages:	25
Keywords:	Methodology, predictions of intentions, vehicles, vulnerable road users

Abstract

This document presents the methodology that will be used in practice to take the steps that will lead to the accurate prediction of intentions of vehicles and vulnerable road users (VRUs). A number of variables have been taken into account in order to describe vehicles intentions, such as road type (highway, local road, urban road), lane width, traffic density in the local vicinity, intersection or round about, inter-vehicle gap, distance to lane marking, lateral velocity, and lateral acceleration. An exhaustive list of potentially discriminant variables has been built accounting for both vehicle variables and contextual information. Similarly, prediction of VRUs intentions and trajectories has also to be validated in an appropriate way, not only in terms of trajectory errors, but also, and more importantly, in terms of action classification delay and criticality of decisions made out of that. For such purpose, a methodological framework has been developed in this task.

Executive summary

One of the most innovative parts of the BRAVE project is to endow vehicles with the capability to communicate with other vehicles that are not equipped with electronic communication devices. This involves the understanding and prediction of manoeuvers of other vehicles. A number of techniques have already been implemented to locate other vehicles on the road, detect their movements, and assess their trajectory. BRAVE builds upon them to estimate the next position of each vehicle and road user sharing the road with the ego-vehicle. Different sensing technologies are currently being used, primarily based on artificial vision, laser, and radar sensors. Data fusion techniques are also necessary in order to further refine the trajectory estimation. Modifications and addition to the vehicle prototype currently available to UAH have also been considered in order to provide a larger and extended sensing coverage. The main goal is to predict the intentions of vehicles and Vulnerable Road Users (VRU) around the ego-vehicle with a view to plan more advanced and safe manoeuvres that account for future movements of other road agents. For this purpose, this document presents the methodology that will be used in practice to take the steps that will lead to the proposed objective.

A number of variables have been taken into account in order to describe vehicles intentions, such as road type (highway, local road, urban road), lane width, traffic density in the local vicinity, intersection or round about, inter-vehicle gap, distance to lane marking, lateral velocity, lateral acceleration. An exhaustive list of potentially discriminant variables has been built accounting for both vehicle variables and contextual information. As a first step, parameters settings and evaluation setup have been defined in a reproducible, comparable manner. Assessment of vehicles prediction of intentions will be carried out on the basis of a manually labeled data-set in which the time-to-action (TTA) moments will be labelled by a human operator as reference ground truth. For each analyzed sequence, the RMSE (Root Mean Square Error) will be computed in a frame range that will comprise half a second before TTA and one second after it (-0.5s, +1.0s). Both lateral and longitudinal errors will be taken into account in the validation phase. As a reference baseline, Kalman filtering will be considered to provide a basic prediction value. Similarly, prediction of VRUs intentions and trajectories has also to be validated in an appropriate way, not only in terms of trajectory errors, but also, and more importantly, in terms of action classification delay and criticality of decisions made out of that. For such purpose, a methodological framework has been developed in this task. As in the case of vehicles, parameters settings and evaluation setup for VRUs have been defined in a reproducible, comparable manner. Thus, average gait cycles of a given number of frames have been considered for different pedestrians. Assessment of pedestrian path prediction will be carried out on the basis of a manually labeled data-set in which the time-to-stop (TTS) and/or time-to-curb (TTC) moments will be provided by a human operator as reference ground truth. For each analyzed sequence, the RMSE (Root Mean Square Error) will be computed in a frame range that will comprise half a second before TTS/TTC and one second after it (-0.5s, +1.0s). Predictions will be evaluated up to 1s ahead in time (compared to ground truth trajectories). Both lateral and longitudinal errors will be taken into account in the validation phase. As a reference baseline, Kalman filtering will be considered to provide a basic prediction value. Regarding pedestrians action classification, the ground-truth data will be manually provided by human operators on the basis of video sequences. Computation of the action detection delay will be considered as a key parameter indicator. For each analyzed sequence, parameters such as accuracy of prediction, precision, and recall of action classification will be computed, as well as ROC graphics. Comparison with human labeled results will be carried out, as a baseline reference for the ceiling of the automatic action classification system.

A key element in the methodological framework is the definition of appropriate use cases, both for vehicles and VRUs, in order to set up a reference action in the testing phase. For that purpose, use cases have been classified into two different categories: vehicle-related, denoted as VEH, and VRU-related (VRU stands for Vulnerable Road Users), denoted as VRU. VEH uses cases will be conducted on month 24 and, if necessary, repeated and perfectioned on month 30. VRU use cases will be tested on month 30. All use cases, VEH and VRU, will be repeated on month 36 on autonomous driving mode. All experiments will be conducted with AUTODRIVE, the autonomous vehicle of the University of Alcalá. This document also describes the different experiments and use cases that will be carried out during the experimentation phase on months 24, 30, and 36, respectively. The experiments provided constitute a selection of the possible use cases proposed by EURO NCAP, as described in the BRAVE proposal.

Document Information

IST Project Number	723021	**Acronym**	BRAVE
Full Title	BRidging gaps for the adoption of Automated VEhicles		
Project URL	http://www.brave-project.eu/		
EU Project Officer			

Deliverable	**Number**	D4.1	**Title**	Methodology for Vehicles and VRUs prediction of intentions
Work Package	**Number**	WP4	**Title**	Vehicle-environment interaction concepts, enhancing ADAS

Date of Delivery	**Contractual**	M9	**Actual**	M9
Status	version 1.0			
Nature	Report			
Dissemination level	Public			

Authors (Partner)	University of Alcalá (UAH)			
Responsible Author	**Name**	Miguel Ángel Sotelo	**E-mail**	miguel.sotelo@uah.es
	Partner	UAH	**Phone**	+34 91 885 6573

Abstract (for dissemination)	One of the most innovative parts of the BRAVE project is to endow vehicles with the capability to communicate with other vehicles that are not equipped with electronic communication devices. This involves the understanding and prediction of manoeuvers of other vehicles. A number of techniques have already been implemented to locate other vehicles on the road, detect their movements, and assess their trajectory. BRAVE builds upon them to estimate the next position of each vehicle and road user sharing the road with the ego-vehicle. Different sensing technologies are currently being used, primarily based on artificial vision, laser, and radar sensors. Data fusion techniques are also necessary in order to further refine the trajectory estimation. Modifications and addition to the vehicle prototype currently available to UAH have also been considered in order to provide a larger and extended sensing coverage. The main goal is to predict the intentions of vehicles and Vulnerable Road Users (VRU) around the ego-vehicle with a view to plan more advanced and safe manoeuvres that account for future movements of other road agents. For this purpose, this document presents the methodology that will be used in practice to take the steps that will lead to the proposed objective.
Keywords	Methodology, predictions of intentions, vehicles, vulnerable road users

Version Log			
Issue Date	**Rev. No.**	**Author**	**Change**
16th January, 2018	0.1	Miguel Ángel Sotelo	Initial version
26th January, 2018	0.2	Miguel Ángel Sotelo, Ignacio Parra, David F. Llorca, J. Alonso, C. Salinas	Internal contributions at UAH
12th February, 2018	1.0	Miguel Ángel Sotelo, Ignacio Parra, David F. Llorca, J. Alonso, C. Salinas	Contributions from BRAVE partners

Table of Contents

Abbreviations

AEB:	Automated Emergency Braking
ADAS:	Advanced Driver Assistance Systems
CNN:	Convolutional Neural Network
DRIVERTIVE:	DRIVER-less cooperaTIVE vehicle
GPLV:	Gaussian Process with Latent Variable
GPDM:	Gaussian Process with Dynamical Model
RMSE:	Root Mean Square Error
ROC:	Receiver Operational Curve
TTA:	Time-To-Action
TTC:	Time-To-Curb
TTE:	Time-To-Event
TTS:	Time-To-STOP
V2V:	Vehicle-to-Vehicle
V2VRU:	Vehicle-to-Vulnerable Road User
VEH:	Vehicle-related use case
VRU:	Vulnerable Road User

1 Introduction

BRAVE addresses the enhancement of ADAS for improving the ‘cooperation’ between vehicles and VRUs. For that purpose, new methodologies are needed in order to accurately estimate the probability that other road users (VRUs or other car drivers) will act unexpectedly by quickly assessing their pose, driving style, speed, position, as well as vehicle’s appearance and status, and even reactions to various possible stimuli generated by the ego-vehicle, through a synergetic effort to advance the on-board perception layer thanks to an unparalleled capability of understanding other vehicles’ manoeuvers and pedestrians’ movements and intentions. The results will offer benefits to automated cars, but also to human-driven vehicles when used in driver assistance systems, ranging from selecting more conservative behaviours when adjacent traffic is aggressive, adapting to the local average driving style in a given geographical area (i.e. increasing other users acceptance), with the ultimate goal of pushing road safety towards new, unmatched levels. When moving in a dynamic environment, the ego-vehicle utilizes a model to predict VRUs and other vehicles’ trajectories, which is based on assumptions defined by traffic rules and confirmed by common sense. However, the expectation is that VRUs and human drivers behave unexpectedly and even break traffic rules making it extremely difficult for the ego-vehicle to plan safe manoeuvres. BRAVE will provide a pioneer solution in an advanced research area close to yielding market products. The consortium work starts on an automated vehicle (DRIVERTIVE) with a limited capacity to interact in a predictive manner with other vehicles and VRUs. By means of BRAVE, new concepts for vehicles and VRUs prediction of intentions will be integrated, providing the possibility to carry out demonstrations in urban traffic conditions. Robust detection and predictive models will be developed in work-package 4 in an attempt to avoid the most typical accidents between vehicles and VRUs, and achieving a high degree of human-like cooperation between vehicles, as well as between vehicles and VRUs. The following picture provides a graphical representation of the different processes and variables involved in this work package, as well as the interactions among them.

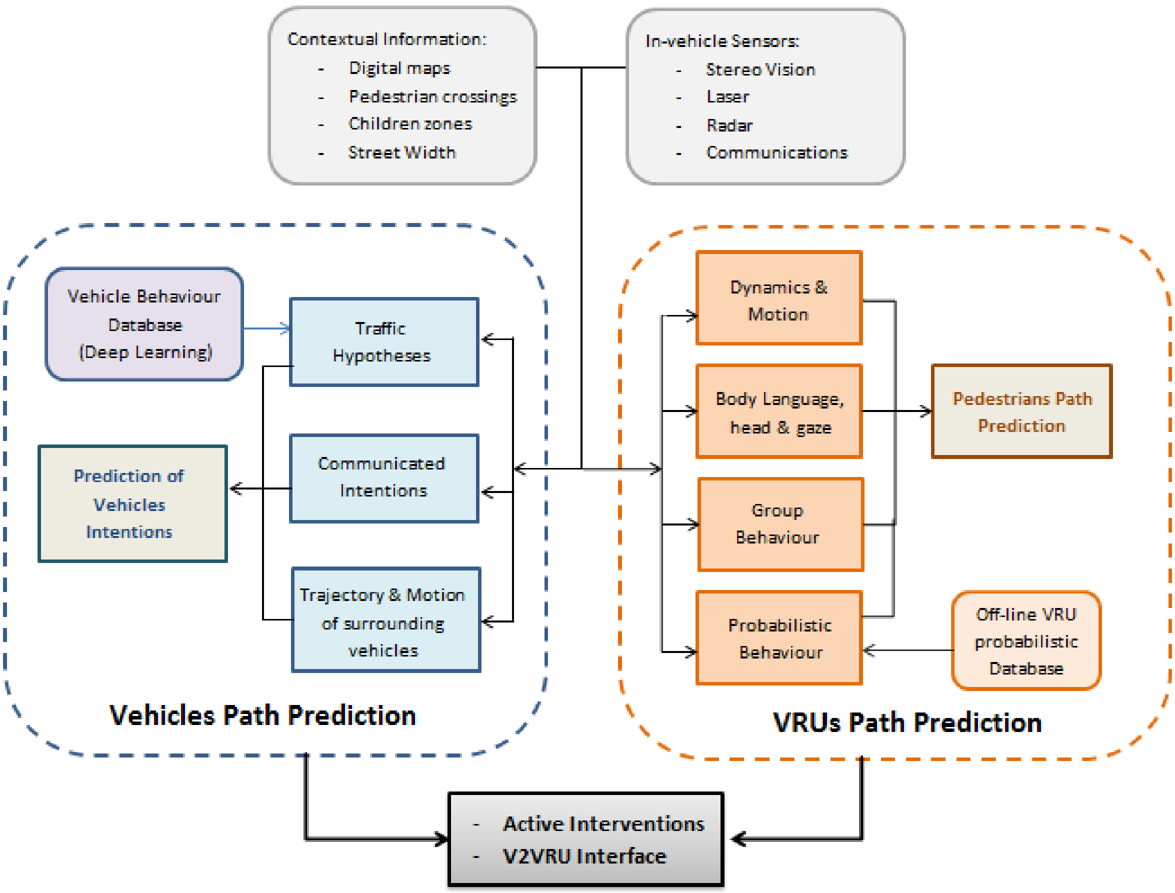

Fig. 1. Structure of processes and variables for the prediction of intentions of Vehicles and VRUs in BRAVE.

One of the most innovative parts of the BRAVE project is to endow vehicles with the capability to communicate with other vehicles that are not equipped with electronic communication devices. This involves the understanding and prediction of manoeuvers of other vehicles. A number of techniques have already been implemented to locate other vehicles on the road, detect their movements, and assess their trajectory. BRAVE builds upon them to estimate the next position of each vehicle and road user sharing the road with the ego-vehicle. Different sensing technologies are currently being used, primarily based on artificial vision, laser, and radar sensors. Data fusion techniques are also necessary in order to further refine the trajectory estimation. Modifications and addition to the vehicle prototype currently available to UAH have also been considered in order to provide a larger and extended sensing coverage. The main goal is to predict the intentions of vehicles and Vulnerable Road Users (VRU) around the ego-vehicle with a view to plan more advanced and safe manoeuvres that account for future movements of other road agents. For this purpose, this document presents the methodology that will be used in practice to take the steps that will lead to the proposed objective. An effective interaction with other traffic participants is an open challenge for intelligent vehicles. This is particularly true in urban environments that are not primarily dedicated to traffic and are populated with Vulnerable Road Users (VRUs) like pedestrians and cyclists. In order to cope with the wide variations in traffic situations and behaviours of traffic participants, scientific progress is required in perception, prediction and interaction techniques.

A number of variables have been taken into account in order to describe vehicles intentions, such as road type (highway, local road, and urban road), lane width, traffic density in the local vicinity, intersection or round about, inter-vehicle gap, distance to lane marking, lateral velocity, and lateral acceleration. An exhaustive list of potentially discriminant variables has been built accounting for both vehicle variables and contextual information. As a first step, parameters settings and evaluation setup have been defined in a reproducible, comparable manner. Assessment of vehicles prediction of intentions will be carried out on the basis of a manually labeled data-set in which the time-to-action (TTA) moments will be labelled by a human operator as reference ground truth. For each analyzed sequence, the RMSE (Root Mean Square Error) will be computed in a frame range that will comprise half a second before TTA and one second after it (-0.5s, +1.0s). Both lateral and longitudinal errors will be taken into account in the validation phase. As a reference baseline, Kalman filtering will be considered to provide a basic prediction value. Similarly, prediction of VRUs intentions and trajectories has also to be validated in an appropriate way, not only in terms of trajectory errors, but also, and more importantly, in terms of action classification delay and criticality of decisions made out of that. For such purpose, a methodological framework has been developed in this task. As in the case of vehicles, parameters settings and evaluation setup for VRUs have been defined in a reproducible, comparable manner. Thus, average gait cycles of a given number of frames have been considered for different pedestrians. Assessment of pedestrian path prediction will be carried out on the basis of a manually labeled data-set in which the time-to-stop (TTS) and/or time-to-curb (TTC) moments will be provided by a human operator as reference ground truth. For each analyzed sequence, the RMSE (Root Mean Square Error) will be computed in a frame range that will comprise half a second before TTS/TTC and one second after it (-0.5s, +1.0s). Predictions will be evaluated up to 1s ahead in time (compared to ground truth trajectories). Both lateral and longitudinal errors will be taken into account in the validation phase. As a reference baseline, Kalman filtering will be considered to provide a basic prediction value. Regarding pedestrians action classification, the ground-truth data will be manually provided by human operators on the basis of video sequences. Computation of the action detection delay will be considered as a key parameter indicator. For each analyzed sequence, parameters such as accuracy of prediction, precision, and recall of action classification will be computed in ROC graphics. Comparison with human labeled results will be carried out, as a baseline reference for the ceiling of the automatic action classification system. A key element in the methodological framework is the definition of appropriate use cases, both for vehicles and VRUs, in order to set up a reference action in the testing phase. For that purpose, use cases have been classified into two different categories: vehicle-related, denoted as VEH and VRU-related (VRU stands for Vulnerable Road Users), denoted as VRU. VEH uses cases will be conducted on month 24 and, if necessary, repeated and perfectioned on month 30. VRU use cases will be tested on month 30. All use cases, VEH and VRU will be repeated on month 36 on autonomous driving mode. All experiments will be conducted with AUTODRIVE, the autonomous vehicle of the University of Alcalá. This document also describes the different experiments and use cases that will be carried out during the experimentation phase on months 24, 30, and 36, respectively. The experiments provided constitute a selection of the possible use cases proposed by EURO NCAP, as described in the BRAVE proposal.

2 Definition of variables for prediction of intentions

Predicting the intentions of VRUs and vehicles requires the measurement of a number of variables of different nature. Some of these variables are context-related while other variables are directly related to the status or appearance of vehicles and VRUs. In this section, a clear listing and definition of the different variables for prediction of intentions is provided. In the case of VRUs the goal is to predict the intentions of pedestrians and cyclists, i.e. starting to cross the street or change trajectory. In the case of vehicles, the intentions to predict have to do with lane change, cut-in or merging manoeuvers, among others, being trajectory prediction a key variable to accurately estimate.

2.1 Variables for prediction of Vehicles' intentions

When predicting the intentions of vehicles in the surrounding of the ego-vehicle a number of factors have to be taken into account. Those factors have to do with the conditions of the environment, such as the road type, number and type of lanes, while other factors deal with observed kinematic or dynamic variables, such as relative speed or acceleration. Thus, the variables involved in the prediction process are divided into two categories: context-related variables and vehicle-related variables.

2.1.1 Context-related variables for vehicle prediction of intentions

The context-related variables have been chosen carefully in order to account for different scenarios on which drivers can make substantially different decisions. The selected variables are described in the following lines.

- **Driving Scenario:** the possible values for this variable are: lane, intersection, roundabout. Driving conditions are very different in these three scenarios, thus drivers decisions may differ significantly. **Road Type:** different types of road can be considered, such as highway, rural road, urban road, etc. Driver decisions for lane change may differ depending on the road type.
- **Lane Width:** this is another important parameter that can affect drivers' decisions on whether or not initiating an overtaking or lane change manoeuver.
- **Traffic density:** the average traffic flow on a given road can create a feeling of safety or discomfort on drivers. This feeling has the potential to affect their decisions when manoeuvring.
- **Number of lanes:** number of lanes in the direction of travel.

The priority in BRAVE is to provide advance prediction of intentions and manoeuvers of vehicles on highways.

2.1.2 Vehicle-related variables for vehicle prediction of intentions

The vehicle-related variables account for kinematic and dynamic conditions. However, other potentially dangerous situations have also been considered in this section, such as, for example, those in which the preceding vehicle has its trajectory abruptly cut-in by other vehicle. In such situation, there is a large probability that the preceding vehicles decide to change lane or at least to decrease velocity significantly and quickly. The selected variables are described next.

- **Longitudinal velocity preceding vehicle:** longitudinal velocity of the preceding vehicle along the ego lane.
- **Longitudinal acceleration preceding vehicle:** longitudinal acceleration of the preceding vehicle along the ego lane.
- **Inter-vehicle gap ego-lane:** this variable measures the relative distance between the ego-vehicle and the preceding vehicle along the ego-lane.

- **Inter-vehicle gap adjoining-lane:** this variable measures the relative longitudinal distance between the ego-vehicle and the closest vehicle along the adjoining-lane (left or right).
- **Relative velocity ego-lane:** relative longitudinal velocity between the ego-vehicle and the preceding vehicle along the ego-lane.
- **Relative velocity adjoining lane:** relative longitudinal velocity between the ego-vehicle and the closest vehicle along the adjoining lane (left or right).
- **Relative acceleration ego**-lane: relative longitudinal acceleration between the ego-vehicle and the preceding vehicle along the ego-lane.
- **Lateral velocity ego-lane:** lateral velocity of preceding vehicle along the ego-lane.
- **Lateral velocity adjoining-lane:** lateral velocity of closest vehicle along the adjoining-lane.
- **Lateral acceleration ego-lane:** lateral acceleration of preceding vehicle along the ego-lane.
- **Lateral acceleration adjoining-lane:** lateral acceleration of closest vehicle along the adjoining-lane.
- **Distance to lane marking:** lateral distance between the vehicle wheels and the closest lane markings. This variable can be applied to the ego-vehicle, to the preceding vehicles along the ego-lane, or to the closes vehicles along the adjoining lanes.
- **Free adjoining lane (left or right):** the adjoining lane (left or right) has enough free space for the vehicles on the ego-lane to safely perform a lane change (left or right). This variable can be applied to the ego-vehicle, to other preceding vehicles along the ego-lane or to other vehicles on adjoining lanes. The possible values for this variable are: true, false.
- **Preceding vehicle approaching a slower vehicle:** this variable indicates that the preceding vehicle is approaching a much slower vehicle. It is a Boolean variable. Consequently, the values that it can take are: true, false. This variable can be applied to preceding vehicles along the ego-lane or to other vehicles along adjoining lanes (left or right).
- **Trailing vehicle on adjoining lane approaching fast:** this variable indicates that the vehicle driving on the adjoining lane (left or right) is approaching fast the ego-vehicle, while coming from behind. In this situation, the ego-vehicle might not have enough time to perform a lane change manoeuver. This is a Boolean variable.
- **Ego-vehicle performing lane change:** this Boolean variable indicates that the ego-vehicle is currently performing a lane change (either to the left or to the right). This situation has the potential to affect the manoeuvring decisions of other drivers in the vicinity of the ego-vehicle.
- **Preceding vehicle performing lane change:** this Boolean variable can be applied to the preceding vehicle along the ego-lane or to other vehicles on adjoining lanes. It indicates that the vehicle under consideration is currently performing a lane-change. This situation has the potential to affect the manoeuvring decisions of other drivers in the vicinity of such vehicle.
- **Communicated intentions:** this variable indicates the intentions of vehicles in the vicinity. These intentions deal with the following actions: left-turn at intersection, right-turn at intersection, continue straight along the same lane at intersection. Communicated intentions are received via V2V.

This is an exhaustive list of variables that can be useful in the intentions prediction process. Nonetheless, it is possible that not all the variables listed on this section will be eventually used in the final model that will be built in the BRAVE project. The final selection of variables will be made after the results obtained during experimentation phase.

2.2 Variables for prediction of VRUs' intentions

Following a similar methodology as in the case of vehicles, the prediction of pedestrians and cyclists in the surrounding of the ego-vehicle involves a number of factors that have to be taken into account. Those factors have to do with the conditions of the environment, while other factors deal with observed kinematic or dynamic variables of the VRUs. Thus, the variables involved in the prediction process are divided into two categories: Context-related variables and VRU-related variables.

2.2.1 Context-related variables for VRU prediction of intentions

As in the case of vehicles, a number of context-related variables have been selected taking into account their potential to influence the decisions of VRUs when it comes to crossing the street. The selected variables are described in the following lines.

- **Scenario:** this categorical variable describes contextual situation, having the possibility to take the following values:
    - Pedestrian crossing: there is a pedestrian crossing in front of the ego-vehicle.
    - Intersection: the ego-vehicle is entering an intersection.
    - Children zone: the ego-vehicle is currently driving on a children zone.
    - School area: the ego-vehicle is currently driving on a school area.
- **Street width:** this variable provides the width of the street on which the ego-vehicle is located.
- **Number of lanes:** number of lanes of the street on which the ego-vehicle is located.
- **Traffic density on the street:** traffic density on the street on which the ego-vehicle is located.
- **Pedestrian density on the street:** pedestrian density on the street on which the ego-vehicle is located.

In order to provide appropriate values for these variables, the ego-vehicle will need access to digital maps and to traffic information. As in the case of vehicles, not all of the listed variables will be necessarily included in the final prediction model. It will depend on availability of information and performance.

2.2.2 VRU-related variables for VRU prediction of intentions

VRU-related variables account for kinematic and dynamic conditions, but also consider other potentially dangerous situations, such as, for example, those dealing with group behaviour. The selected variables are described next.

- **VRU body language:** this is a very informative variable, especially when it comes to pedestrians. The possible values that it can take are the following:
    - Pedestrian walking: pedestrian performing a regular walking action.
    - Pedestrian standing: pedestrian standing on the sidewalk.
    - Pedestrian starting: pedestrian is transitioning from standing to walking.
    - Pedestrian stopping: pedestrian is transitioning from walking to standing.
    - Cyclist straight: cyclist is pedalling on straight pose.
    - Cyclist bending: cyclist is bending (inclining his/her body) to the left or to the right. This is a clear indication that the cyclist is turning or about to turn (to the left or to the right).
- **Gaze direction:** this variable indicates whether or not the pedestrian or cyclist is turning his/her head towards the oncoming traffic in order to look for eye contact with the driver. In such a case, it

provides a clear indication of the intention of the VRU (to cross the street, in the case of pedestrians, or to turn, in the case of cyclists).

- **Cyclist turn indication:** this variable indicates whether or not a cyclist is raising his/her arm to indicate his/her intention to turn (either to the left or to the right).
- **Group behaviour:** the behaviour of other VRUs has the potential to strongly influence the behaviour of other VRUs in the vicinity. Thus, this variable indicates the average behaviour of other pedestrians, as a group. For example, if a group of three pedestrians are crossing the street chances are that a fourth pedestrian, walking behind them, takes the decision to start crossing the street, even if the safety conditions to do that are not properly met.
- **Relative distance to curb:** relative distance between a given VRU and the closest curb.
- **Relative distance to oncoming vehicle:** relative distance between a given VRU and the closest oncoming car.
- **Relative velocity to cyclist:** relative longitudinal velocity between a preceding cyclist and the ego-vehicle.

3 Parameters setting and KPI

The key parameters indicators (KPI) must be adequately defined in order to provide an evaluation framework that allows to reproduce and compare results in a proper manner. Assessment of vehicles and VRUs prediction of intentions will be carried out on the basis of a manually labelled data-set that will be used as a reference ground-truth. For that purpose, a number of variables are defined as indicators of key moments in the prediction timeframe. Those variables are described below.

- **TTA (Time-To-Action):** this variable indicates the remaining time from the current time until the time when the next action will take place. This action can be one of the following:
  - A car starts a lane change manoeuver.
  - A car starts a merging manoeuver.
  - A pedestrian starts to walk.
  - A cyclist starts a turning manoeuver.
- **TTS (Time-To-Stop):** this variable indicates the remaining time until the moment in which a pedestrian comes to a full stop.
- **TTC (Time-To-Curb):** this variable indicates the remaining time until the moment in which a pedestrian steps on the road curb.

In the case of pedestrians, TTA and TTS can be considered as a whole indicator denoted as TTE (Time-To-Event). These indicators provide the key variables to undertake the performance analysis of the prediction systems to be developed in BRAVE.

3.1 Performance indicators for Vehicles' prediction of intentions

The analysis of performance will rely on two main indicators when it comes to predicting the intentions and manoeuvers of vehicles. Their description is provided below.

- **Trajectory error:** the real trajectory carried out by the vehicle under analysis will be compared with the predicted trajectory for such vehicle in a given time interval (up to 4s). For each analysed sequence, the RMSE (Root Mean Square Error) will be computed in a frame range that will comprise half a second before TTA and one second after it (-0.5s, +1.0s). Both lateral and longitudinal errors will be taken into account in the validation phase. As a reference baseline, Kalman filtering will be considered to provide a basic prediction value.
- **Detection delay:** it is defined as the difference between the starting time of a real event (lane change or lane merging manoeuver) and the time when the system predicts such event. The difference can be positive, meaning that the detection provided by the prediction systems is lagging, or negative, indicating that the prediction system has some anticipation capability based on contextual information. As a reference baseline, manually labelled data will be used for performance comparison.
- **Event recognition:** this variable measures the performance exhibited by the prediction system to recognize all possible events (lane changes, lane merging). Given that it is possible that the system triggers some false detection, a proper performance analysis will be conducted using a ROC that will combine Precision and Recall indicators. In addition, the F-1 measure (based on precision and Recall) will be computed. As a reference baseline, manually labelled data will be used for performance comparison.

3.2 Performance indicators for VRUs' prediction of intentions

Similarly, prediction of VRUs intentions and trajectories has also to be validated in an appropriate way, not only in terms of trajectory errors, but also, and more importantly, in terms of action classification, detection delay and criticality of decisions made out of that. As in the case of vehicles, parameters settings and evaluation setup for VRUs will be defined in a reproducible, comparable manner. Thus, average gait or pedalling cycles of a given number of frames will be considered for different VRUs. Assessment of VRU path prediction will be carried out on the basis of a manually labelled data-set in which the time-to-stop (TTS) and/or time-to-curb (TTC) moments will be provided by a human operator as reference ground truth. The following variables are defined for the sake of performance analysis.

- **VRU Trajectory error:** the real trajectory carried out by the VRU under analysis will be compared with the predicted trajectory for such VRU in a given time interval (up to 1s). For each analysed sequence, the RMSE (Root Mean Square Error) will be computed in a frame range that will comprise half a second before the event and one second after it (-0.5s, +1.0s). Both lateral and longitudinal errors will be taken into account in the validation phase. As a reference baseline, Kalman filtering will be considered to provide a basic prediction value.
- **VRU Detection delay:** it is defined as the difference between the starting time of a real event (starts to walk, starts to turn, etc.) and the time when the system predicts such event. The difference can be positive, meaning that the detection provided by the prediction systems is lagging, or negative, indicating that the prediction system has some anticipation capability based on contextual information. As a reference baseline, manually labelled data will be used for performance comparison.
- **VRU Activity recognition:** this variable measures the performance exhibited by the prediction system to recognize all possible activities (walking, stopping, pedalling, turning, etc.). Given that it is possible that the system triggers some false detection, a proper performance analysis will be conducted using a ROC that will combine Precision and Recall indicators. In addition, the F-1 measure (based on precision and Recall) will be computed. As a reference baseline, manually labelled data will be used for performance comparison.

In addition to the quantitative performance assessment variables described in this section, a qualitative assessment of performance will be carried out based on the criticality of decisions made by the prediction system.

4 Use-cases definition

BRAVE will test a number of use cases related to vehicles and vulnerable road users, mainly pedestrians, as described by the recommendations issued by EURO NCAP. A selection of those use cases will be implemented and tested during the experimentation phase of the project. In all cases, the selected use cases deal with anticipated vehicle behaviour in order to enhance safety when interacting with other vehicles or with VRUs (pedestrians and cyclists). The EURO NCAP uses cases are devised for testing systems aiming to enhance current ADAS or even to provide advanced operation of self-driving cars in complex situations. The criteria used to select the use cases to be tested is based on the requirement of increasing user acceptance of self-driving cars by exhibiting advanced functionality beyond the current state-of-the-art. The description of the different use cases for vehicles and VRUs, together with the preliminary planning for conducting the experimentation phase is provided in the following sections.

4.1 Vehicle-related use cases

In this section, a preliminary description of the selected EURO NCAP use cases for interaction with other vehicles is provided. These use cases consider complex situations on highways and will be emulated on proving grounds in France and Spain.

- **Use Case VEH-1**: a car enters the highway from the ramp aggressively while the ego-vehicle drives along the right-most lane, as graphically depicted in figure 2. The ego-vehicle must be able to sense the entering vehicle using some of the on-board sensors, such as radar and laser, or receive the intentions of the oncoming vehicle by means of a communications link, and to react appropriately by giving way, change lane or whatever that is deemed appropriate as a function of the situation (relative velocity and acceleration, intersecting trajectories, safety gap, etc.). The ego-vehicle will continue without changing its speed or without performing any lane change manoeuver in case there is no conflict between its predicted trajectory and the predicted trajectory of the vehicle entering the highway.

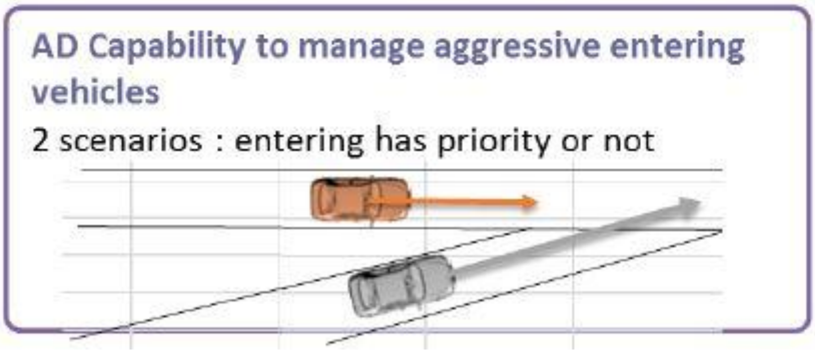

Fig. 2. Graphical description of Use Case VEH-1.

- **Use Case VEH-2**: a preceding car is changing lane with no prior signalling while infringing the safety gap with respect to the ego-vehicle (for example, due to sudden difficult perception prior to entering a tunnel or approaching a change of slope). The relative velocity between the lane-changing car and its predecessor along its lane is very high. The adjoining lane of the lane-changing car has sufficient free space. All the conditions for a lane change manoeuver are met. Those conditions should be anticipated by the ego-vehicle, so that it can react in an anticipated manner (for example, by decreasing velocity or by changing lane if possible in order to create a safe situation for all cars involved in the scene). Figure 3 shows the case of a sudden change in perception conditions caused by a tunnel, but a similar situation could be considered for a change of road slope.

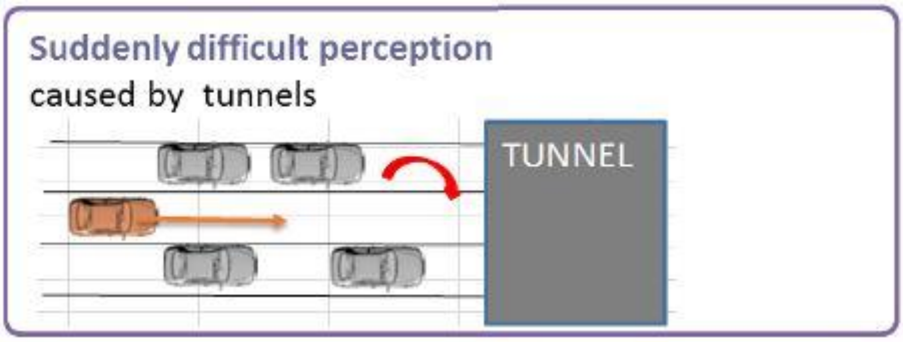

Fig. 3. Graphical description of Use Case VEH-2.

4.2 VRU-related use cases

In this section, a preliminary description of the selected EURO NCAP use cases for interaction with VRUs is provided. These use cases consider complex situations on urban and road scenarios, as depicted in figure 4.

Euro NCAP 2016& 2018 VRU rating scenarios in discussion (UTAC & Euro CAP WG's)		Speed range VUT	VRU	Test parameters	KSI	rating Euro NCAP
AEB VRU - Pedestrian 2016 & 2018	Crossing - Current scenarios	20-60 km/h	5 & 8 km/h	Impact location: 25-50-75% child obstruction		Euro NCAP 2016 4 tests
	Longitudinal scenario	20-80 km/h	5 km/h	Impact location: 25% , 50%	15%	Euro NCAP 2018
	Darkness	20-60 km/h	5 km/h	Impact location: 25% , 50% , 75%	58% fatalities	Euro NCAP 2018 ? TBC
AEB VRU - Cyclists 2018	Crossing from both sides	20-60 km/h	15-20 km/h	with / without obstruction	54% Killed 57% SI	Euro NCAP 2018
	Turning	10-40 km/h	15-20 km/h		2% Killed 5% SI	Euro NCAP 2020
	Longitudinal	40-80 km/h	15-20 km/h	AEB (lower speed) & FCW (higher speed) impact location 50-80%	24% Killed 7% SI	Euro NCAP 2018

Fig. 4. Graphical description of Use Cases for VRUs, as defined by EURONCAP.

As in the case of the vehicle-related use cases, VRU-related use cases will be tested on proving grounds in France and Spain using dummies. The selected use cases for testing in BRAVE are described in the following lines.

- **Use Case VRU-1**: a pedestrian steps onto the street, continues to walk and cuts the ego-vehicle trajectory. The ego-vehicle must be able to detect the crossing intention in an anticipated manner and act accordingly. The car will decrease speed significantly (coming to a full stop if necessary) and signal the pedestrian by switching on the GRAIL interface.
  - Similar performance should be attained if the system detects that the pedestrian is keeping eye contact with the driver.
  - The car should decrease velocity in a preventive manner if the system detects that a pedestrian is walking along the sidewalk in parallel to the road, even if no intention to cross is detected (but there are chances that it might happen).

- **Use Case VRU-2:** a pedestrian crosses the street all of a sudden intersecting the ego-vehicle's trajectory, creating a dangerous situation for the pedestrian and for the car's occupants. The car should be able to perform an automatic emergency braking (AEB) manoeuver. Only if it is safe enough, the car could perform an avoidance manoeuver.

- **Use Case VRU-3:** the system detects a cyclist in front of the ego-vehicle. The car should overtake the cyclist while maintaining an appropriate safety distance (lateral distance) and reducing the speed accordingly. The overtaking manoeuver must be performed only if there is free space along the ego-lane and the adjoining lane (no oncoming traffic). Otherwise, the car must reduce speed and stays behind the cyclist until the conditions for a safe overtake are met.

5 Creation of Data-set

As in any other Machine Learning problem, the need for data is one of the major issues. Consequently, a large data-set will be built in the course of the BRAVE project. This data-set will comprise thousands of examples of vehicles and VRUs performing a large variety of actions in different contexts. The main goal is to provide the data for learning but also to provide the means for accurate and solid assessment of performance. There are lots of data-sets available, but none of them exhibits the characteristics that are needed for intention prediction purpose. Hence the need to create a new data-set that fully fits the needs of the problem at hand. UAH will develop a brand-new data-set from scratch comprising Vehicles and VRUs (Pedestrians and Cyclists). The final data-set will be made public to the scientific community with a view to become a milestone in the field.

5.1 Data-set features

The main features that this data-set will exhibit are the following:

- Multi-object class: the data-set will contain several object classes, namely Vehicles, Pedestrians, and Cyclists.
- Multi-instance: all individual instances of the different objects present on the scene will be labelled separately.
- Kinematic, Dynamic, and Contextual Information: the data-set will contain information of different nature, such as kinematic, dynamic, and context-related (as described in section 2).
- Fully labelled at frame level: all frames in the data-set will be labelled, not only a number of key frames per second.
- Object tracklets: the full object tracklets, comprising the entire trajectory of each object while visible on the scene, will be made available.
- Event and Activity information: information about events and activity (for VRUs) will be labelled so that indexed search will be possible. Thus, users of the data-set can easily locate lane change manoeuvers, for example.
- High accuracy labelling: labelling of each object will be done in a highly accurate manner.

In order to ease the creation of the data-set, a semi-automatic tool will be developed by UAH. This tool will enable the manual labelling of the initial instance of a given object, while using advanced tracking techniques to provide automatic tracking and location on subsequent frames. Similarly, CNNs will be used to mask the input image in order to better localize the appearance of relevant objects in the scene. This tool will significantly ease the labelling task.

5.2 Event-labelling methodology

Event labelling is a complex task that can derive in subjective results and significant differences among labelling operators. This problem becomes especially relevant when it comes to labelling transition events. Labelling transitions for vehicles and cyclists is left to the assessment of manual operators. The transitions (lane change or lane merging for vehicles; turning manoeuver for cyclists) will be labelled on the basis of video sequences. The transition will happen when the vehicle or the cyclist starts to change their status. For such purpose, operators will label the frame in which such change of status starts. They will label the transition to the best of their expertise. Given that operators will have the possibility to move forward and backward along the video sequence, no major variance on the labelled data is expected. However, labelling of transitions for pedestrians is much more variable. Accordingly, this section provides basic guidelines for easing the event-labelling task for pedestrians and to produce as much robust results as possible.

This guideline allows identifying the instant when a pedestrian starts or finishes performing a given action, such as starting or stopping. Thereby, the proper definition of the different events and activities is provided in the following lines.

- Starting activity: it is defined as the action that begins when the pedestrian moves one knee to initiate the gait and ends when the foot of that leg touches the ground again.
- Stopping activity: it is defined as the action that begins when a foot is raised for the last step (stride) and finishes when that foot touches the ground again.

Examples of transitions manually labelled from standing to starting, starting to walking, walking to stopping and stopping to standing are shown in Figures 5, 6, 7, and 8 respectively. In these pictures, pedestrians are represented schematically by using dots that represent the different body parts and joints (head, neck, shoulders, knees, etc.). By following the criterion established in this document, pedestrian-related activities can be easily labelled by human experts, thus enabling the creation of reliable ground-truth data.

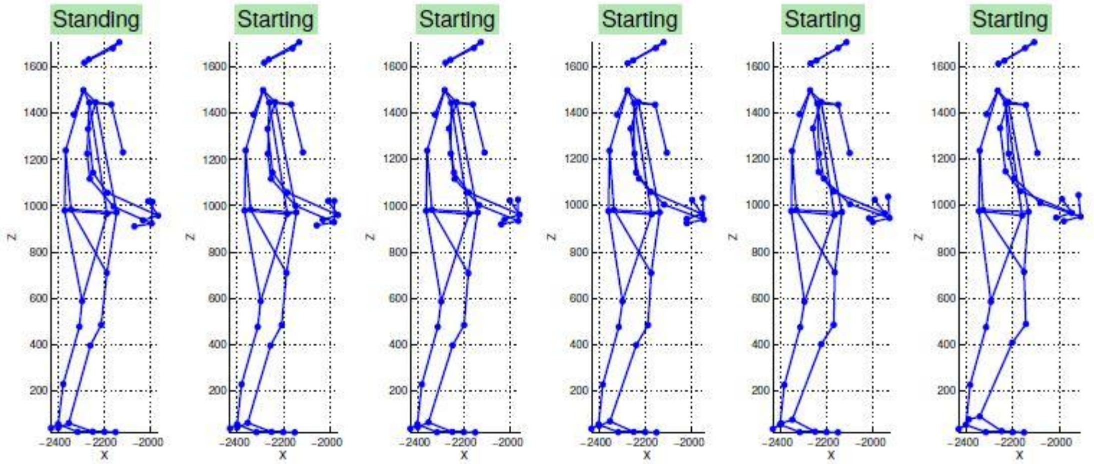

Fig. 5. Manually labelled transition: Standing-to-Starting.

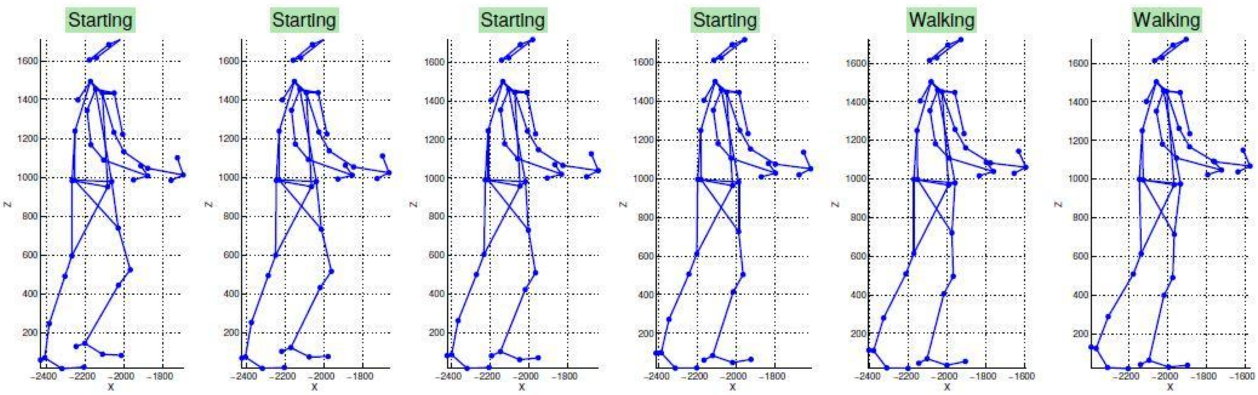

Fig. 6. Manually labelled transition: Starting-to-Walking.

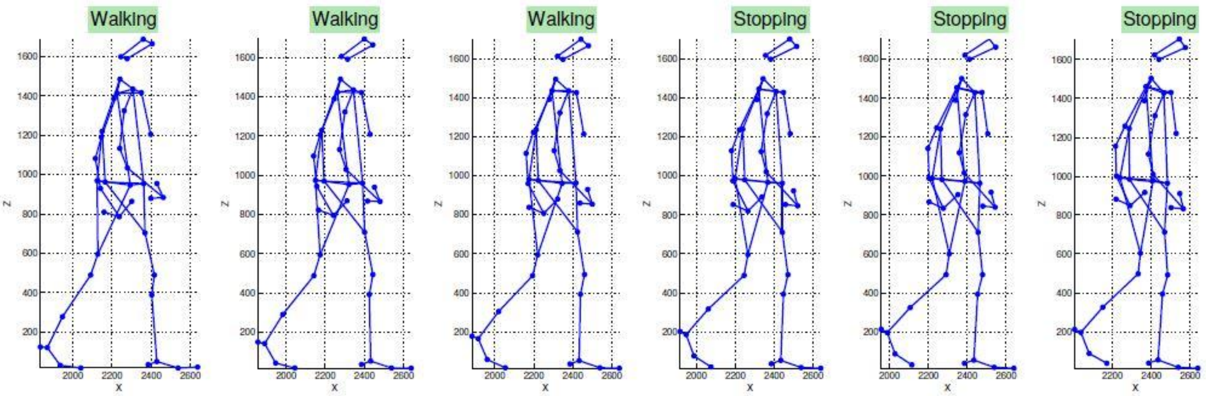

Fig. 7. Manually labelled transition: Walking-to-Stopping.

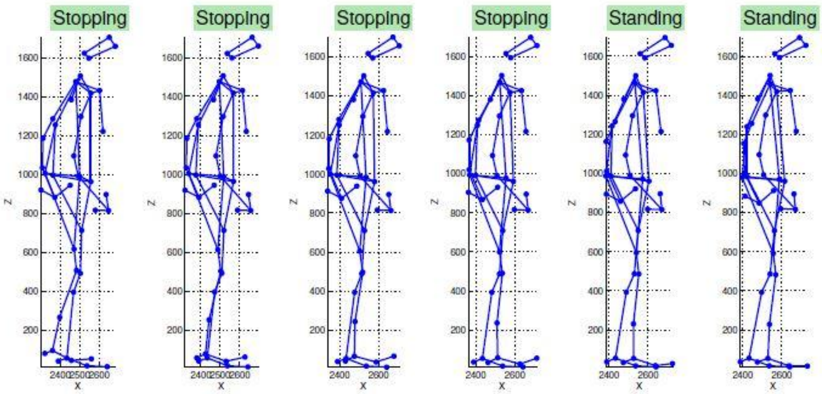

Fig. 8. Manually labelled transition: Stopping-to-Standing.

5.3 Sensing requirements

The creation of the data-set will be possible by using the following sensors installed on-board DRIVERTIVE UAH automated vehicle.

- Vehicle data:
  - Velodyne 32, forward looking colour camera, rear looking colour camera.
  - Accurate calibration among these three sensors will be developed in order to provide all-around sensing in a range of 100 m around the ego-vehicle.
- VRU data:
  - Monocular colour vision (forward looking) and radar.
  - Calibration between vision and radar system will be developed.

6 Learning methodology

This section describes the methodology that will be followed in order to develop learning systems capable of anticipating the intentions and most likely trajectories of vehicles and VRUs. The purpose of this document is not to provide detailed technical descriptions of the systems and algorithms to develop, but to provide a clear description of the different methodological steps that must be taken in order to build prediction systems for vehicles and VRUs.

6.1 Learning methodology for VRUs

In order to learn the most likely intentions and trajectories of VRUs (pedestrians and cyclists), the following methodological steps are considered. The detailed technical development and description of the different steps will be the goal of future deliverables.

- **VRU skeleton estimation.** The main goal is to locate the presence of VRUs in the forward scene using a forward looking vision system. The bounding box provided by the vision system can provide the main input for that purpose. Once VRUs are properly located a detailed analysis and identification of their body parts and joints must be carried out in order to adequately and individually locate their heads, shoulders, arms, legs, knees, etc. These individual body parts and joints will provide the main clue for dynamic prediction of pedestrians and cyclists. The use of Deep Learning techniques (CNNs) is envisaged to facilitate the detection of such parts in a robust manner based on the colour image provided by one of the two forward looking cameras.
- **VRU motion learning**. Once the VRU body parts and joints are located, a temporal sequence of such elements must be considered (for example, the past 1-2 seconds acquired by the vision system) in order to build the input to the motion learning system. The motion learning system will identify the VRU action (walking, standing, stopping, starting, pedalling, swerving left or right, etc.) and will learn the most likely VRU trajectory in a time horizon of 1 second ahead, based on the prior knowledge of VRUs dynamics. For this purpose, the motion learning process will be carried out off-line. Different techniques will be tested and compared, such as GPLV (Gaussian Process with Latent Variable), GPDM (Gaussian Process with Dynamical Model) and CNNs (Convolutional Neural Networks).
- **VRU trajectory estimation**. This process is not part of the learning process itself. It must be run online in real time while the ego-vehicle is acquiring data from on-board sensors. The systems must be able to provide the most likely VRU trajectory in a time horizon of 1 second ahead using the knowledge gathered during the learning phase. The same algorithms tested during such phase will be compared in practice in terms of accuracy and RMSE of predicted trajectory.
- **Learning VRU intentions.** The final decision on VRU intentions must be build based on several factors: VRU trajectory estimation (based on motion learning), group behaviour, contextual situation, VRU behaviour (looking for eye contact with driver, turning head towards oncoming traffic, etc.). All these variables will be gathered using the forward looking cameras (group behaviour, VRU behaviour, and motion prediction) and a prior digital map (contextual situation) and applied as input to a probabilistic graphical model that will provide the final intentions of all VRUs in the scene. The use of Bayesian Networks is envisaged in this learning phase although other techniques might also be considered.

6.2 Learning methodology for Vehicles

Following a similar approach, the following methodological steps are considered in the learning process for predicting the trajectories and intentions of vehicles. The detailed technical development and description of the different steps will be the goal of future deliverables.

- **Vehicle location.** The main goal is to locate the presence of vehicles on the scene. Given that vehicles can appear at any position around the ego-vehicle, the system will analyse the images provided by the forward and rear looking cameras as well as the information provided by the Velodyne-32. The use of deep learning techniques is envisaged for the case of vision-based data, while point-cloud analysis tools will be needed for the case of data provided by the Velodyne-32. Vision systems are expected to be useful in the range of 100 meters in front or behind the ego-vehicle. Velodyne-32 is expected to provide data for accurate vehicle localization in a range of 40 meters around the ego-vehicle. Appropriate correspondence of vision-laser data will be carried out in order to enhance vehicle detections and to avoid duplicated detections.

- **Vehicle trajectory estimation.** This process is a purely kinematic and dynamic estimation of the most likely trajectories for all vehicles around the ego-vehicle in a time horizon of 4 seconds ahead. The result of this process constitutes a valuable input to the situation analysis function. This estimation will be carried out using traditional machine learning algorithms and CNNs. Comparison among them will be provided.
- **Situation analysis.** The purpose of this phase of the learning process is to correctly interpret the situational constraints that affect the ego-vehicle and other vehicles around it. For example, a vehicle on the adjoining lane that is quickly approaching a much slower vehicle on the same lane is very likely to perform a lane change manoeuver in order to overtake such slow vehicle. The lane change manoeuver might turn out in a cut-in action for the ego-vehicle. This is a very clear example of the kind of situations that must be learnt and inferred by the situation analysis module.
- **Learning vehicles intentions.** The final decision on vehicles intentions must be built based on several factors: vehicle trajectory estimation, situation analysis, prior behaviour knowledge. All these variables will be gathered using the forward and rear looking cameras, the Velodyne-32, and a prior digital map (contextual situation) and applied as input to a probabilistic graphical model that will provide the final intentions of all vehicles in the scene. As in the case of VRUs, the use of Bayesian Networks is envisaged in this learning phase although other techniques might also be considered.

7 Evaluation and testing methodology for intentions prediction

The full experimentation and validation plan of BRAVE is part of WP5. However, those aspects dealing with assessing the performance of prediction systems are briefly presented in this section in order to make this document self-contained.

7.1 Evaluation methodology for VRU intentions prediction

The evaluation of performance for this system will be carried out following the key parameters indicators and key variables that were described in section 2. For this purpose, the following phases are considered.

Assessment of Activity Recognition Performance

The system will be tested on different conditions using video sequences containing pedestrians and cyclists performing a variety of actions covering all the possible VRU behaviours considered in this project. The video sequences will be recorded in the following scenarios:

- UAH University Campus and city of Alcalá de Henares (Madrid, Spain).
- UTAC (France).

Assessment of performance will be carried out with a data-set containing at least 500 different VRU activities. A minimum performance of 90% correct detection rate will be sought. Similarly, the detection delay will be analysed in term of average, standard deviation, median, minimum, and maximum values. A target value is set at 150ms of average detection delay in order to match or even better the typical human delay response.

Assessment of VRU path prediction results

The quality of the trajectory prediction system for VRUs will be carried out by computing the combined lateral and longitudinal errors at different TTE (Time-To-Event) times, from -1 to +1 s. At each TTE considered the predicted trajectory will be analysed at different time horizons, from 0.25s to 1.0s. The system will be tested with at least 1.000 different video sequences.

7.2 Evaluation methodology for Vehicle intentions prediction

The evaluation of performance for this system will be carried out following the key parameters indicators and key variables that were described in section 2. For this purpose, the following phases are considered.

Assessment of Activity Recognition Performance

The system will be tested on different conditions using video sequences containing vehicles performing a variety of actions covering all the possible vehicle behaviours (lane change, cut-in, etc.) considered in this project. The video sequences will be recorded in the following scenarios:

- Alcalá de Henares, Madrid, and Guadalajara (Spain).
- UTAC (France).
- Vransko (Slovenia).

Assessment of performance will be carried out with a data-set containing at least 200 different vehicle activities. A minimum performance of 90% correct detection rate will be sought. Similarly, the detection delay will be analysed in term of average, standard deviation, median, minimum, and maximum values. A target value is set at 200ms of average detection delay in order to match or even better the typical human delay response.

Assessment of Vehicle path prediction results

The quality of the trajectory prediction system for vehicles will be carried out by computing the combined lateral and longitudinal errors at different TTE (Time-To-Event) times, from -1 to +1 s. At each TTE considered the predicted trajectory will be analysed at different time horizons, from 1.0 to 4.0s. The system will be tested with at least 100 different video sequences.

7.3 Preliminary timing

The preliminary timing for conducting the different experiments and testing the use cases is provided in the following lines.

- Month 24. Uses cases VEH-1 and VEH-2 will be tested on manual driving mode (ADAS mode). All tests will be conducted at the premises of UTAC (France).
- Month 30. Uses cases VRU-1, VRU-2, and VRU-3 will be tested on manual driving mode (ADAS mode) at the premises of UTAC (France).
- Month 36. All use cases will be repeated with real users on autonomous mode at the premises of ACASA (Spain).

8 Final remarks

This document has presented the methodology for learning the predictions of intentions of Vehicles and VRUs. For such purpose, different KPIs and key variables have been defined in order to set a standard for performance assessment. The following remarks are highlighted regarding the content of this document.

- This methodological document describes the most likely techniques and methods to be used during the project to the best of our knowledge. Given that the state-of-the-art of mathematical techniques for machine learning and prediction is making progress very quickly, it is possible that some other techniques, different from those mentioned in this document, are used in the final prediction systems to be developed in the framework of BRAVE.
- The international scientific community is very active in terms of data-set creation. Thus, it is foreseeable that some new data-set for prediction of intentions will appear in the course of the BRAVE project. If that were the case, we envisage to incorporate the data provided by such data-set in BRAVE evaluation and testing phase. This strategy would definitely add value to the developments carried out by the BRAVE team.