

## AN INTELLIGENT CO-DRIVER SURVEILLANCE SYSTEM

MĂDĂLINA TOMA, MIRELA POPA, LEON ROTHKRANTZ<sup>a, b, \*</sup>

<sup>a</sup> *Intelligent Interaction, Delft University of Technology, Mekelweg 4, Delft, The Netherlands*

<sup>b</sup> *Faculty of Transportation Sciences, Czech Technical University in Prague, Konviktská 20, Prague 1, Czech Republic*

\* corresponding author: [L.J.M.Rothkrantz@tudelft.nl](mailto:L.J.M.Rothkrantz@tudelft.nl)

**ABSTRACT.** In recent years many car manufacturers developed digital co-drivers, which are able to monitor the driving behaviour of a car. Sensors in the car measure if a car passes speed limits, leaves its lane, or violates other traffic rules. A new generation of co-drivers is based on sensors in the car which are able to monitor the driver behaviour. Driving a car is a sequence of actions. In case a driver doesn't show one of the actions the co-driver generates a warning signal. Experiments in the car simulator TORC were performed to extract the actions of a car driver. These actions were used to develop probabilistic models of the driving behaviour. A prototype of a warning system has been developed and tested in the car simulator. The experiments and test results will be reported in this paper.

**KEYWORDS:** behaviour analysis, car simulator, surveillance, Bayesian reasoning.

---

### 1. INTRODUCTION

In recent years we can observe an increase of investments of car producers and researchers in the development of self-driving cars. The first prototypes of such cars have been tested. However, there is a long way to go before such cars appear on the highways and in the cities. It can be expected that at least for the coming 10 years cars with a human driver will prevail. Also in the regular driver-car interaction a lot of automatization will take place. Navigation systems support the driver to find the shortest route from his starting point to its destination. Smart interfaces support the drivers to use their phones, radio and other digital devices in their cars. This paper focuses on the development of systems increasing the safety of the car driver and other road users.

At Delft University of Technology there is a project running on the development of surveillance angels and guardian angels [1], [2]. One of the applications is a digital automated car driver assistant. Such an assistant is able to supervise the driving behaviour of a car driver. The system permanently receives the position of the steering wheel, position of the pedals, gear lever, but also information of body movements of the driver, its facial expression and speech and information from the environment. This information is provided by sensors in the car and sensors attached to car parts and (manual-) instruments. The goal of the system is to make assessments of the possible dispositions/situation of the car driver, that is to say a semantic interpretation of the driving behaviour. In case the driver overlooks important information, violates traffic rules or driving rules, the system will generate an alert or support the driver. The next step is that the car takes over the driving of the car, however this is postponed to future work.

The full model is based on information of the car driver, but also on the environment and state of the car. From the movements of the car driver we make assessments of the state of the car driver, his goals and intentions. Information on the state of the driver can also be assessed by physiological measures and EEG measurements [3],[4]. This study focuses on the assessment of the visual behaviour of the car driver. A car driver can show many body movements. We can order them in specific categories:

- *Core driving movements.* These movements are related to driving a car.
- *Peripheral movements.* These movements are related to interaction with peripheral devices in the car such as radio, phone, navigation devices.
- *Signal movements.* These movements express some message from the driver about his physical or emotional state (stretching the arms, showing a fist) or messages to other drivers on the road.
- *Random movements.* These movements have no specific semantic meaning related to car driving such as scratching the nose.

Core driving movements can be composed of one single action or a sequence of actions. Take, for example, the actions of the car driver such as starting a car, parking, speeding up, slowing down, taking a turn. All these actions are characterized by a sequence of information of the car sensors. After an action the next step is to classify the action. It is necessary to realise that the classification is context sensitive. The classification is usual an ambiguous probabilistic process. In case of an ongoing sequence of actions we have to predict the future and compute the probability of possible next steps. It can also happen that a driver

skips one or more action on purpose or by mistake. Take, for example, the case when a driver wants to take over but does not view the mirror to check if the next lane is free. This can result in a dangerous situation and an alert has to be generated. In this paper a rule based system and a Bayesian reasoning system will be used to compute the probability of missing or future steps in an action string. By experiments using a driving simulator all possible sequences of actions are computed as well as corresponding probabilities of those actions. In the next section we discuss related work. In section 3 a model of a digital co-driver is presented. In section 4, experiments with the car driver simulator TORC are discussed. Then we present an analysis of the recorded data in section 5. This paper is concluded with a final discussion.

## 2. RELATED WORK

According to the statistics published by the National Highway Traffic Safety Administration in 2005 a big number of collisions (78%) and near-collisions (65%) are associated with driver inattention. Distractions caused by secondary tasks, such as manipulation of the navigation system, radio, or phone, seems to be the main source of the inattention [5]. In [6] Merat and Jamson present their findings regarding driver's ability to respond to sudden and unexpected events while driving under normal conditions and also while performing other tasks. Their conclusions highlight that the reaction time increases with 200ms when the driver is using in-car systems. Head pose estimation reveals important information about the driver's focus of attention. In [7] Doshi and Trivedi present their findings regarding head dynamics and eye gaze as important cues in predicting driver's intent of changing lanes.

Regarding human pose estimation, currently there are various representation methods of a human in an image, such as silhouettes [8], bounding boxes, or sticks representing the limbs of a person [9][10]. One of the most popular approaches is using the pictorial structures model which stands as a collection of human body parts [11].

Building intelligent systems have become a trend in current researches. In driving domain, they have been called advanced driver assistance systems [12]. According to Adarsha [13], these systems support the driver's sensing ability and they are able to track and detect the error/lapses of drivers. Assistive intelligent systems are usual a combination of several tools. For example, Benoit, [14] developed in his work an assistive driving system for assessing the driver's fatigue. His system uses two platforms, namely OpenInterface developed in C++ for signal processing and ICARE. ICARE is a conceptual component model for multimodal input/output interaction. Thus, the Benoit's system combines multimodal signal processing analysis with a multimodal interaction. Our system can be

built on a similar system architecture, and it tries to enhance the novice driver's skills.

Many other intelligent systems have been developed [15]. The cognitive model of the driver had been investigated in order to implement driver's mental activity in this system. The driver's mental activity is analyzed when a driver interacts dynamically with the driving environment. Therefore, the driver mental representation of a situation can be illustrated as an instance in the working memory. We also investigate in building a cognitive model of drivers. We analyzed driver's mental activity as a result of the following activities: gaze and head activities related to the traffic situation, the car controller (steering wheel, pedals, and gear shifter) related to body parts movements, and states of the car on the road.

Other intelligent systems have been implemented based on a stochastic model. In these cases, driver behavior has been modeled using collected data from the driving simulator. Liang used this model in his work [16], to detect driver distraction in real-time. He modeled driver behavior based on parameters extracted from driver's eye movements, states of the steering wheel and car position on the lane. In a similar way Giusti [17] detected driver sleep-attacks by using acquired data from the steering wheel.

Wang and Gao estimated in [18] the state of a car on the road with a rule-based expert system. The expert system typically provides the intelligence of the system, and it fits very well with the requirements of our system. According to the last decade researches, an expert system can be classified in the following six methodologies: rule-based systems, knowledge-based systems, intelligent agent (IA), database methodology, inference engine, system-user interaction. A basic rule-based expert system has the following entities: a knowledge base which contains data from experts, an inference engine which contains the evaluation of decision making components of the system. We estimated the driver's intent with a rule-based expert system. Moreover, our framework developed with a rule-based expert system reasons like a driver providing full assistance for novice drivers in terms of the driving skills enhancement.

Fletcher proposed an assistive driver system [19], which monitors driver activity using vision sensors. The set of sensors used in Fletcher's system tracks the eye movements and the body parts of the driver. However, it is difficult to estimate driver intentions from the data of the visual sensors. Therefore, our system estimates driver's intention by correlating data from visual sensors with states of cars and information about the driving environment. The gaming industry provides driving simulations which simulate 3D traffic environments. If a simulation of a 3D traffic environment is used, the data about a driving situation can be read easily. In our work, we used TORCS simulator because it can be extended and adapted to the proposed system. This simulator is an excellent

open source tool, and it gave us an opportunity to configure the tracking sensors and the car controller.

In prior research, we tried to determine the driver's interaction intent by analysing only single body parts, e.g. pose estimation [20], gaze detection, or facial expression. Compared with these, we take into account more aspects and actions, such as body postures, gaze direction, and head orientation during interaction with the car and the surrounding environment. A lot of intelligent systems that provide assistance to the drivers were presented above. Some of them present good solutions. Still an intelligent system with a complex reason which trains the novice driver's skills need to be developed. Thus, we present the implementation of an intelligent system that has the capability to assist a novice driver and it tries to enhance his skills at the same time.

### 3. MODEL

In figure 1 we display a model/architecture of our digital co-driver system. The behaviour of the driver is assessed by parallel sensor systems: KINECT, gaze tracker, car-driver interaction tools. The sensory input system has been configured and synchronized with clock cycle of TORCS tool in order to track the driver activities. The TORCS simulator is extended on three modular levels of abstraction. In the first level, the system detects sequences of postures. The driver action can be recognized from sequence of postures in the second level. In the last level, a driver's intention is estimated based on the driver's actions related to the traffic situation. The set of possible intentions had been stored in the database. When a driver's intention is wrong the proposed framework sends feedback alarms. A GUI Interface has been built in order to allow monitoring the functionality of the whole system and also driver's activity.

The main reasoning module of the system implements the intelligence of the system, and it is responsible to predict driver's intent. In one of the developed prototypes a rule based system was used to assess the most probable scenario. Every rule has a tally attached to it to indicate the probability or importance of the rule. Some scenarios are composed of one single action. An observed action triggers the most probable rule. This activated rule is allowed to fire and the right hand side of the if-then rule is generated. Some scenarios consist of only one action. If this is a dangerous driving action an alert has been generated by a rule. An example is a driver who doesn't look at the road in front of him. In other cases observed action(s) trigger the most probable scenario and then specific rules check if important actions are missing or dangerous actions can be expected. An example is a car driver switching on his signal lights to change his lane who did not check of the next lane is free to insert. The problem with rule based systems is that all the rules have to be defined in advance. So

the system is not prepared for new, unexpected situations. Another problem is that sensors are not able to assess all the ongoing actions. These errors in the sensor observation can be caused by failing technology or because vision field of sensors is occluded. The actions are ambiguous or the lighting conditions are too bad. A gaze tracker, for example, usually misses a significant amount of actions. By sudden, large movements of the head the gaze tracker has to be calibrated again. Nevertheless, the biggest problem to solve is that actions are usually distributed in time and that some car drivers start actions simultaneously. It is necessary to realize that the proposed system will be used as a decision support system. Ultimately car drivers themselves remain responsible for their actions. To improve the performance of the assessment of the actions a Bayesian probabilistic system will be used. To compute the values of the entries in conditional probabilities tables experts are needed to set these values or huge amount of data generated in experiments testing the system is needed to compute the values. The robustness of the system is increased with a fusion method based on multiple sensors compared with a fusion method based on a single sensor. Multi-sensor data fusion allows the combination of information with different physical characteristics to enhance the understanding of the driver's action. The information read from the set of sensors are fundamental in the decision-making module of the system. Our fusion method seeks to combine information from multiple sensors. Table 1 presents a hierarchy of driver activity in a bottom-up way, and it represents the foundation of our system architecture. The first level needs to detect the driver postures, and then sequences of postures define the driver action. In the third level, one action or a set of actions predict the driver's intention. A human intention can be defined as an anticipated outcome that guides planned actions. It is the goal or purpose behind an action or a set of actions that a person is following. Driver's intention can be predicted from the driving actions. Driving actions in a scenario can be recognized based on a set of postures performed over time. Thus, the system architecture depicted in Figure 1 defines an assistive car driving system based on multi-steps of processing: tracking driver behavior, predicts driver's future behavior, decide if the driver's intention is correct or not, and issue corrective feedback.

### 4. SIMULATION ENVIRONMENT

The main tool in our system is TORCS simulator. It is one of the most popular 3D open source car simulators written in C++ and available under GPL license. It can be used as an ordinary car racing game, as an AI racing game. The main advantage of the software is that we could extend the platform by configuring inside all sensors used for tracking the driver's behavior. We selected TORCS in our research for the following reasons. TORCS is an advanced fully

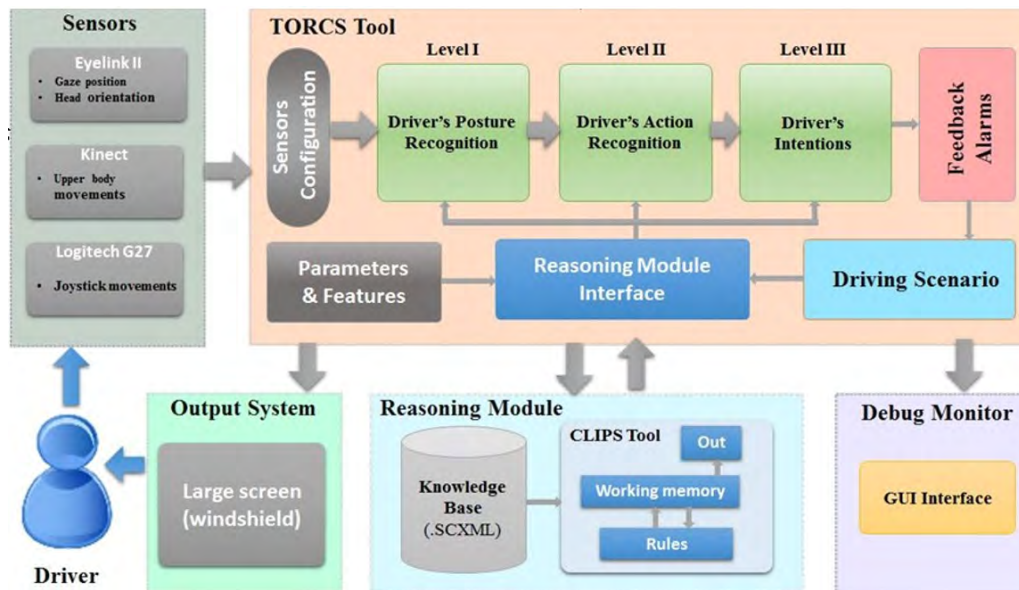


FIGURE 1. Architecture of the digital co-driver system.

Levels of driver activity	Driver interaction with car controller and traffic situations
<div style="display: flex; align-items: center;"> <div style="writing-mode: vertical-rl; transform: rotate(180deg); margin-right: 5px;">Level III</div> </div>	<p>Driving intent interaction: is a mental state of any driver who executes an action or set of actions into a driving situation based on driving environment and the car states.</p>
<div style="display: flex; align-items: center;"> <div style="writing-mode: vertical-rl; transform: rotate(180deg); margin-right: 5px;">Level II</div> </div>	<p>Driver's actions: are made up of multiple body-part gestures such as eyes, head, arms, legs motion, and each body-part gesture is an elementary event of motion and can be composed of a sequence of instantaneous poses at each moment of time.</p>
<div style="display: flex; align-items: center;"> <div style="writing-mode: vertical-rl; transform: rotate(180deg); margin-right: 5px;">Level I</div> </div>	<p>Driver's postures: are instantaneous configuration of the body part in the space of interaction.</p>

TABLE 1. Hierarchy of driver activity.

customizable simulator, and it can be adapted for our application; it features a sophisticated physics engine as well as a 3D graphics engine for the visualization of the virtual environment; and it has a modular software architecture, hence the new controlling and sensing devices allow a straightforward integrating. According to the features of the proposed system, we design tracks which simulate some driving scenarios. Therefore, the creation of a customized track involves the use of other tools. In the first step, we used the Trackeditor tool for creation a track. It can design the tracks adding straight or curves and some parameters can be configured such as length, radius or banking. All of this information can be stored in an XML file. In addition, the Trackeditor tool generates a file with extension. AC, and this file stores the 3D description of the track. The 3D description of the track was edited with the Blender Tool. Blender tool allows us to add elements such as traffic signs and to insert

textures. We added billboards to one of the tracks.

Looking at the driver and understanding his actions is the first task of our system. We used a marker-less system to track upper body activity, and the driver's upper limb motion can be tracked by the Kinect device. This device has been developed by Microsoft as a game console input device. Comparing with other video cameras [21] used for recognizing human's movement activities, it tracks the motion of the subject through a combination of hardware and software technologies and achieves a high accuracy tracking of the body with a rate of 30 FPS (frames per seconds).

In our system, we used for eye movements and head orientation the EyeLink II device produced by the Canadian SR Research, which uses pupil and cornea reflection tracking mode. This system provides a comprehensive tracking based on smart vision sensors which consist of a head-mounted camera system and two PCs for processing data and running experiments.

On the head-mounted device, both left and right eye pupil position and the head orientation relative to computer monitor can be tracked. Combining the position of the head with the pupil movement relative to the screen, enabled recording of the gaze direction. Sensors for physical driver's interaction.

Our proposed system (see figure 2, figure 3 and figure 5) presents the architecture of a Driver-Computer Interaction system. The driver can interact naturally with the simulated traffic environments through physical input/output interaction interfaces by bridging the gap between the digital and physical world.



FIGURE 2. Car driving simulator.



FIGURE 3. Typical scene from the curved race track with billboards.

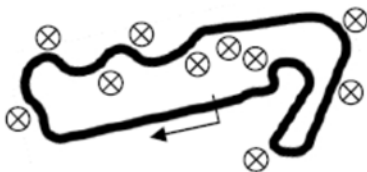


FIGURE 4. The curved race track driven by the participant. Markers indicate locations of the billboards, if enabled.

We used as an input interface a Logitech G27 joystick controller based on a steering wheel, a gear shifter and three pedals (clutch, brake and throttle). It is similar to the usual basic control system from a cockpit car and the driver's actions can be performed in the same way as in the real car.

The driving scenes are displaying on an output interface which is a large screen like a real car windshield.

We used a TV screen with 56 inches size diameter as a large screen. Four optical sensors of head camera from the Eyelink II device had placed in the four corners of the screen.

## 5. EXPERIMENTS

In this section we discuss two experiments testing respondents in our driving simulator TORC as displayed in figure 5. Several devices were used to assess the actions of the driver. In section 5.2 we discuss the assessment procedures using the KINECT device. But first we discuss the experimental results of assessment of driving actions using the car sensors as discussed in section 3.

Data captured with video cameras and gaze tracker has been captured with different sampling rates. The same holds with data from the KINECT sensor. Recorded data is redundant, full of errors and missing data. The different data streams have to be fused. Different types of data fusion using multiple, multimodal streams have been discussed in [22].



FIGURE 5. Test person in action in the driving simulator.

### 5.1. EXPERIMENT 1 DETECTION BEHAVIOUR ACTIONS

In total, there were 23 students invited to take part in the driving experiment. They were supposed to drive for two hours, one hour on a simple trajectory and one hour on a curved trajectory. They had to drive with different speed, normal speed and as fast as possible. Along the simple trajectory we placed some billboards. There were other car drivers on the road driving both in a low and high speed speed. This forced our test persons to take over at regular times. We were interested in isolated driving actions or scenarios composed of single driving actions. In figure 6 we show a state diagram of all possible states of actions and transition between states. During driving, the students got phone calls or visual messages about routing. The billboards along the road caught the car driver's attention.

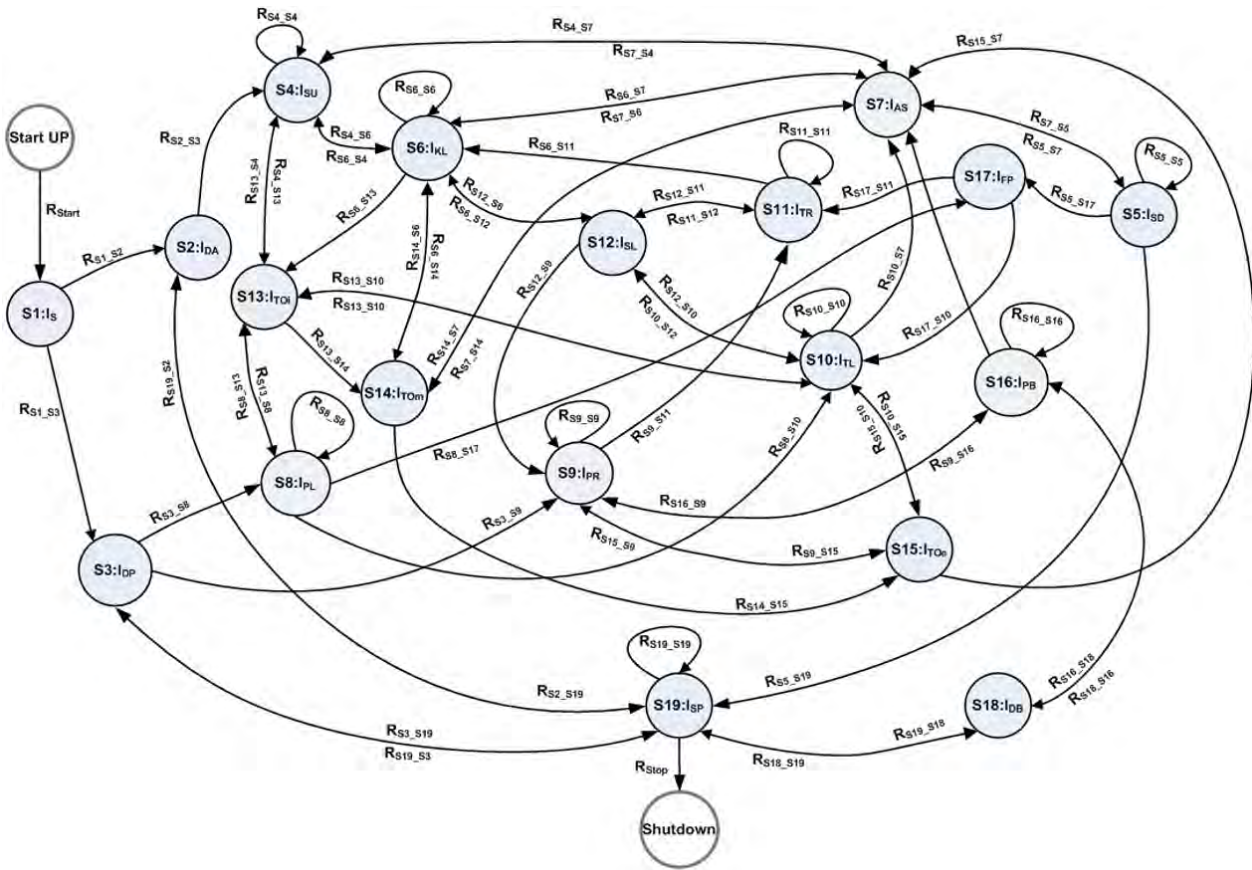


FIGURE 6. The state diagram describing the driver intention and rules used on each transition.

5.1.1. ANALYSIS EXPERIMENT 1

The goal of our system is to generate alerts in case of dangerous driver’s behaviour. We analyse single actions and scenarios composed of actions. A car driver is supposed to watch the road all the time. In case he/she is looking left or right for some time, looking to the billboards of traffic signs for a long time or looking at their telephone or routing devices for a long time, a dangerous situation can occur. Assessing the gaze direction via the gaze tracker is rather complicated. Fast movements over large view angles were difficult to track. Sensors in the car detected if the car left its lane. In the next section we discuss our experiments using KINECT to assess the position of the head and gaze direction.

Parallel to the single action analysis we researched possible scenarios. As soon as the action has been detected we find the position in the state diagram and research if this state is the starting point of a scenario or a point within a scenario. If an important preceding action is missing an alert can be generated. In case of the take-over scenario it can happen that the driver switch on the signal lights but monitoring the side mirror is not detected. To reduce the amount of false alarms we used a probabilistic approach. States in the state diagram and transition between states have probability numbers depending on the frequencies in all recordings during the experiment. We choose the most probable states and transition path to compute

the possibility of an alert.

We tested 23 drivers during 2 hours (high-low speed). On average, every minute a single action or a scenario of average 3 actions of the driver was required or generated. In total we observed 16.500 single actions and 48.230 single action in scenarios (see table 2). We focused on the actions with most frequent errors.

5.2. EXPERIMENT USING KINECT

This section presents one of the first digital co-driver experiments [23]. We present adapted behavioral models, which are based on our driving experience and also on information gathered from experts in the field. We assess both head pose and body pose in terms of orientation and hands position in relation to the car environment. We defined normal vs. dangerous driver’s behavior, on several levels of importance.

Normal driver’s behavior implies both hands on the steering wheel and the head orientation frontal. There are variations from these rules, such as: head orientation can be left or right for a small amount of time in case of lane changing or observation of the surroundings, etc. The dangerous behavior appears in case the driver keeps looking in any other direction than frontal for a long period of time.

Regarding hands position, the driver is supposed to have at least one hand on the steering wheel, while the other one can rest, be used for changing the gears or

Actions/Scenarios	Frequency errors	High speed	Low speed	Correct alert
		Correct Alert	Frequency	
Looking around	121	72%	276	81%
Switching lanes	322	68%	158	77%
Giving priority	220	75%	118	87%
Speed limit	296	92%	88	94%

TABLE 2. Frequency of detected actions generated alerts.

for manipulating the radio or the navigation system. If the driver doesn't keep his hands on the steering wheel for a long period of time, it is considered to be a dangerous action. Furthermore, if besides having no hands on the steering wheel, the driver is also looking in another direction we consider this behavior very dangerous and the system will generate an alert to notice the driver.

A very important aspect taken into consideration while defining the behavioral models is the temporal evolution of an action. Under normal circumstances, stretching arms (no hands on the steering wheel) or looking in another direction, these activities are performed by a driver without any serious consequences. Still, in case any of these actions are performed for a longer period of time they might affect the level of concentration and of the driver's attention. What is more, in case of an unexpected or sudden event, the driver's speed of reaction might not be high enough, which could lead to an accident.

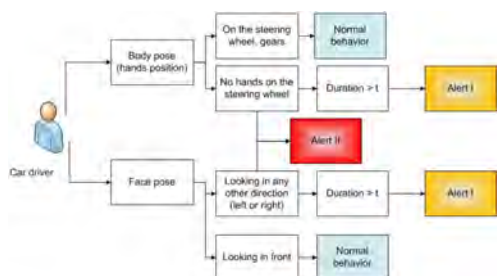


FIGURE 7. Architecture of a digital car driver system.

KINECT software module was used for pose estimation. The output of the pose estimation module, consisting of the location and orientation of each body part (see figure 8) was used to assess the relation between the different body parts (e.g. arms relative to the torso, lower arms relative to upper arms). We computed the angles between the different body parts on a frame basis and also for an image sequence. The angles between head, torso and the vertical axis together with the one between head and torso provide relevant information regarding head and body position. We considered three basic orientations of the head relative to the body: straight, left, and right.

The goal of our research is to test the possibility to train the basic car driver scenarios using a driving simulator and serious gaming. In table 4 we displayed

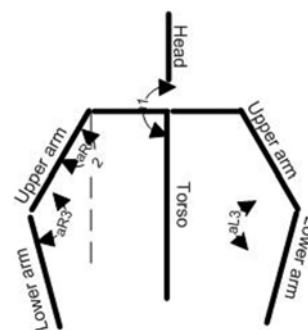


FIGURE 8. Model of the upper body used for pose estimation.



FIGURE 9. Hierarchy of driver activity.

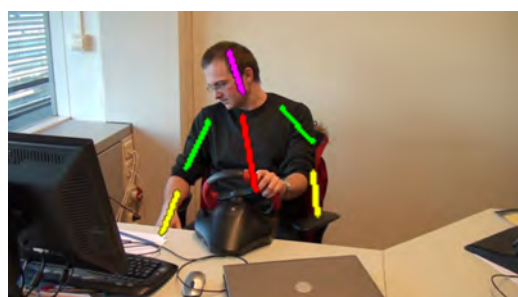


FIGURE 10. Incorrect detection of left arm.

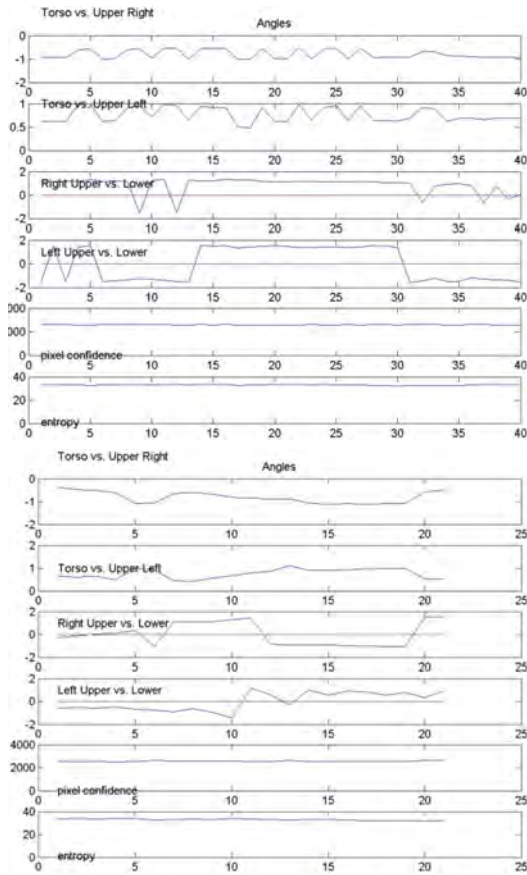


FIGURE 11. Motion angles pattern for two actions: (a) hands on the steering wheel (b) hands raised .

some of the basic scenarios. Every scenario is composed of a sequence of basic actions. We selected the following basic actions:

- Browsing (moving position of head, looking into the mirror, inspecting the routing device or radio)
- Tuning (radio, mobile phone etc.)
- Activating (touching eyes, mouth, pulling ears etc)
- Selecting (putting the gear by hand in the requested position)
- Waving (putting left/right hand up)
- Driving (turning the steering wheel).

We evaluated the performance of the body pose estimation module on the driving recorded data, by inspecting visually the accuracy of each detection (see figure 9 and figure 10), and we achieved 89% correct upper body detections and 58% correct upper body parts estimation.

Kinect proved to be very succesfull with limb detection and person tracking. Another important thing about Kinect's limb tracking is that Kinect is aware of occlusion and for example if one of the arms is missing, the stick configuration shows that a limb is missing. In short two main problems that are faced with Kinect and the video tool are undetected limbs (see table 5) and frames with no detection at all respectively

(see table 6). The detection heavily depends on the exposed action.

In table 3, we show the average missing rate.

The car environment can be divided into a number of regions of interest for the driver. The most important and used ones are the 'steering wheel' (1) and the 'gears' (2) regions, followed closely by the 'contact' (3) and 'brake' (4) regions. Other secondary regions are the 'navigation system' (5), the 'radio' (6) and the 'drawer' (7) ones. A visual representation of the defined regions can be found in Figure 12.



FIGURE 12. Regions of interest inside a car.

We included a graphical division of the inside the car environment into regions of interest. The driver's interaction with each object inside the car can be assessed by determining hands' position inside a region. This information is used in combination with the motion pattern analysis in order to extract a first semantical interpretation of the possible action of the driver.

Movements characterized by a certain speed and amplitude, inside a specific region of interest are very likely to depict a certain type of action. For example movements having a low speed and a small amplitude associated with the 'steering wheel', 'radio' or 'navigation system' regions of interest are very probable to correspond to driving, manipulating the radio or the navigation system actions. On the other hand, movements with a high speed and a large amplitude inside the 'drawer' or the 'gears' regions are associated with picking an item from the drawer or changing the gears actions.

The ROIs, the transitions between the different ROIs, the associated type of movements, and the probable behavioural interpretation are presented in Figure 13.

The reasoning step was implemented using a rule-based system which received as input, features from the previously described modalities (body pose estimation, face detection, and regions of interest assessment). On a time basis the extracted features were observed and using the rules contained in the state-based model we were able to distinguish normal from potential dangerous and dangerous driving behavior. Conclusion regarding the driving behavior was generated. We employed the state-based model depicted in Figure 13.



	Browsing	Tuning	Activating	Selecting	Waving	Driving
Browsing	23	22	55	0	0	0
Tuning	0	95	5	0	0	0
Activating	5	10	77	3	0	5
Selecting	0	17	0	83	0	0
Waving	0	0	0	14	85	0
Driving	25	0	25	0	0	50

TABLE 3. Confusion matrix of selected driving actions.

Driver action	Behavior detected by KINECT, video and car sensors
<b>S1 Starting the car</b>	
1. Check if the gear is in neutral position	Hand movements,
Movements of head (turn down)	
2. Start the car	Hand motion
3. Press the throttle a little bit	Movements of right leg
<b>S2 Driving away</b>	
1. Start the car	See S1
2. Press the clutch to the floor	Movement of left leg
3. Enter the gear shift in position 1 (movement of the head)	Hand movement
4. Press the throttle pedal	Movement of right leg
5. Release the clutch	Movement of left leg
<b>S3 Driving away from parked position</b>	
1. Start the car	See S1
2. Look in the inner and outer mirror if next lane is free	Turning the head right/left
3. Switch on the left signalling light	Movement of hand
4. Press the clutch to the floor	Movement of left leg
5. Enter gear shift in position 1	Movement of hand,
Movement of leg	
6. Press slowly the throttle pedal	Movement of right leg
7. Turn the steering wheel to the left or right	Hand movement
8. Turn the steering wheel to the neutral position after reaching the next lane	Hand movement

TABLE 4. Scenarios and behavioural cues.

		Total number of frames
KINECT	16%	574
Video-tool	32%	574

TABLE 5. Rate of frames without detected limbs.

	Upper right arm	Upper left arm	Lower right arm	Lower left arm
Browsing	12	74	20	74
Tuning	10	2	12	2
Activating	2	0	2	0
Selecting	12	7	12	7
Waving	16	75	21	75
Driving	17	20	17	20

TABLE 6. Kinect rate of frames without detected limb relative to all frames in the dataset.

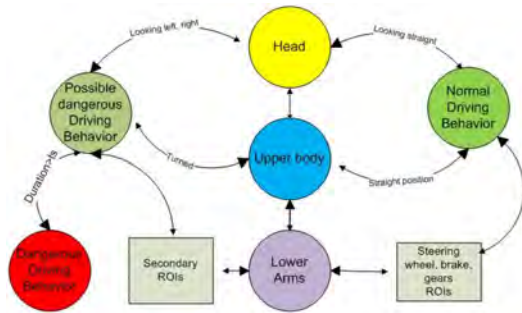


FIGURE 13. State based model.

We used 3 seconds as a threshold for the critical duration of a secondary action or body part orientation, such as head orientation, lower arms position, or body pose orientation.

### 6. CONCLUSIONS

In the framework of designing a digital co-driver system, we performed a study to assess the actions of a driver. Based on the experiments in a car simulator it was shown that the action of the car driver can be assessed in more than 80% of the cases. But we stress the fact that the results are based on the experiments in a car simulator using students as testpersons doing their best to drive according to the rules. In real life situations the results could be quite different.

We expect that in real life situations the recognition rate could be quite lower because of bad lighting conditions in the car, occlusions, position of the car driver. That is the reason that we used multiple multimodal sensors. This implies sensor fusion. To assess the driver actions we used a probabilistic rule based system. The limitation of such a system is that all possible (string of) actions have to be defined in advance. Unexpected rare actions will not be detected. This is also caused by the fact that our system selects the most probable actions. In case that our system has to detect some actions with a high priority to send an alert such actions have to be labeled with a higher priority in our system.

The designed system can be used as a decision support system in current cars. The car has to be equipped with some sensors and a microprocessor to process the data and generating alerts. The results can also be used in the design of self driving cars

in the near future by modelling a digital car driver according to a human model.

The basic positions of individual body parts, along with the motion temporal patterns and the corresponding regions of interest were fused in order to draw a conclusion regarding the driver’s possible type of activity.

### REFERENCES

- [1] L. Rothkrantz. Surveillance angels. *Neural Network World* **24**:1–25, 2014.
- [2] L. Rothkrantz. Smart surveillance systems, network topology. In *Command and Control: Organization, Operation, and Evolution: Organization, Operation, and Evolution*, pp. 270–290. 2014.
- [3] P. Bouchner, M. Hajný, S. Novotný, et al. Car simulation and virtual environments for investigation of driver behavior. In *Proceedings of the 7th WSEAS International Conference on Automatic Control, Modeling and Simulation, ACMOS’05*, pp. 523–530. World Scientific and Engineering Academy and Society (WSEAS), Stevens Point, Wisconsin, USA, 2005.
- [4] M. Haak, S. Bos, S. Panic, L. Rothkrantz. Detecting stress using eye blinking and brain activity from eeg. In *Proceedings of the 1st Driver Car Interaction and Interface, DCII’08*, pp. 35–60. 2008.
- [5] V. Neale, T. Dingus, S. Klauer, et al. An overview of the 100-car naturalistic study and findings, national highway traffic safety administration.
- [6] N. Merat, A. H. Jamson. Multisensory signal detection: How does driving and ivis management affect performance? In *Proceedings of the 4th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, p. 351–357. 2007.
- [7] A. Doshi, M. M. Trivedi. On the roles of eye gaze and head dynamics in predicting driver’s intent to change lanes. *Trans Intell Transport Sys* **10**(3):453–462, 2009. DOI:10.1109/TITS.2009.2026675.
- [8] V. Ferrari, M. Martin-Jimenez, A. Zisserman. 2d human pose estimation in tv shows. In *Proceedings of the Dagstuhl Seminar on Statistical and Geometrical Approaches to Visual Motion Analysis*. 2009.
- [9] P. Bouchner. Car simulation and virtual environments for investigation of driver behavior. *Neural network World* **15**(2):149–163, 2005.

- [10] M. Andriluka, S. Roth, B. Schiele. Pictorial structures revisited: People detection and articulated pose estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2009.
- [11] H. I., H. D., D. L. S. Ghost Ū a human body part labeling system using silhouettes. In *Proceedings of the 14th International Conference on Pattern Recognition*. 1998.
- [12] H. Winner, S. Hakuli, F. Lotz, C. Singer (eds.). *Handbook of Driver Assistance Systems: Basic Information, Components and Systems for Active Safety and Comfort*. Springer, Cham, 2016. DOI:10.1007/978-3-319-12352-3.
- [13] R. Adarsha, V. Kumar, K. Ganesan. Low cost driving trainer assistance system. *Journal of Transportation Technologies* **2**(1):63–66, 2012. DOI:10.4236/jtts.2012.21007.
- [14] A. Benoit, L. Bonnaud, A. Caplier, et al. Multimodal signal processing and interaction for a driving simulator: Component-based architecture. *Journal on Multimodal User Interfaces* **1**(1):49–58, 2007. DOI:10.1007/BF02884432.
- [15] T. Bellet, B. Bailly-Asuni, P. Mayenobe, A. Banet. A theoretical and methodological framework for studying and modelling driversŠ mental representations. *Safety Science* **47**(9):1205 – 1221, 2009. Research in Ergonomic Psychology in the Transportation Field in France, DOI:http://dx.doi.org/10.1016/j.ssci.2009.03.014.
- [16] Y. Liang, M. L. Reyes, J. D. Lee. Real-time detection of driver cognitive distraction using support vector machines. *IEEE Transactions on Intelligent Transportation Systems* **8**(2):340–350, 2007. DOI:10.1109/TITS.2007.895298.
- [17] A. Giusti, C. Zocchi, A. Rovetta. A noninvasive system for evaluating driver vigilance level examining both physiological and mechanical data. *Trans Intell Transport Sys* **10**(1):127–134, 2009. DOI:10.1109/TITS.2008.2011707.
- [18] J.-H. Wang, Y. Gao. Multi-sensor data fusion for land vehicle attitude estimation using a fuzzy expert system. *Data Science Journal* **4**(1):127–139, 2005.
- [19] L. Fletcher, L. Petersson, A. Zelinsky. Driver assistance systems based on vision in and out of vehicles. In *IEEE Proceedings. Intelligent Vehicles Symposium, 2003*, pp. 322–327. IEEE, 2003.
- [20] M. M. Trivedi, S. Y. Cheng, E. M. C. Childers, S. J. Krotosky. Occupant posture analysis with stereo and thermal infrared video: algorithms and experimental evaluation. *IEEE Transactions on Vehicular Technology* **53**(6):1698–1712, 2004.
- [21] J. Shotton, A. Fitzgibbon, M. Cook, A. Blake. Real-time human pose recognition in parts from single depth images. In *Proceedings of the CVPR*. 2011.
- [22] I. Lefter, L. J. M. Rothkrantz, G. J. Burghouts. A comparative study on automatic audio-visual fusion for aggression detection using meta-information. *Pattern Recogn Lett* **34**(15):1953–1963, 2013. DOI:10.1016/j.patrec.2013.01.002.
- [23] M. Popa, L. Rothkrantz. Assessment of behaviour in serious games of driving simulators. *International Journal of Intelligent Games & Simulation* **6**(2), 2011.