

AFFECTIVE COMPUTING AND AUGMENTED REALITY FOR CAR DRIVING SIMULATORS

DRAGOȘ DATCU^{a,*}, LEON ROTHKRANTZ^{b,c}

^a *Twnkls, Groothandelsgebouw, Stationsplein 45, Rotterdam, The Netherlands*

^b *Intelligent Interaction, Delft University of Technology, Mekelweg 4, Delft, The Netherlands*

^c *Faculty of Transportation Sciences, Czech Technical University in Prague, Konviktská 20, Prague 1, Czech Republic*

* corresponding author: email@dragosdatcu.eu

ABSTRACT. Car simulators are essential for training and for analyzing the behavior, the responses and the performance of the driver. Augmented Reality (AR) is the technology that enables virtual images to be overlaid on views of the real world. Affective Computing (AC) is the technology that helps reading emotions by means of computer systems [1][2][3], by analyzing body gestures, facial expressions, speech and physiological signals. The key aspect of the research relies on investigating novel interfaces that help building situational awareness and emotional awareness, to enable affect-driven remote collaboration in AR for car driving simulators. The problem addressed relates to the question about how to build situational awareness (using AR technology) and emotional awareness (by AC technology), and how to integrate these two distinct technologies [4], into a unique affective framework for training, in a car driving simulator.

KEYWORDS: Affective Computing, Augmented Reality, Serious Gaming, Car Driving Simulator.

1. INTRODUCTION

Due to its capability to improve perception of reality, to support collaboration, a visual display of virtual objects, and to enable transitions between real and virtual environments, augmented reality (AR) technology can be used to create novel interfaces for face-to-face and remote collaboration for training [5][6].

Affective computing-based technologies for car driving simulators have capability to lower the cost of accessing the expertise (by reducing a need to move trainers to where their expertise is needed), and to increase availability of expertise. This applies primarily to multi station driving simulators. AR technology can support new types of visualization and can help to develop new learning experience for the trainee. The whole driving simulation can be rendered using AR, using AR glasses, screens or windshield projections.

By using augmented reality, the trainee can be presented various stimuli taking the form of virtual representations. In addition, visual notifications can be presented, as generated automatically by the system or as specifically instructed by expert. In addition to augmented reality, reading driver's affect by affective computing technology has capability to adapt the simulation given the affect state of the trainee.

1.1. AUGMENTED REALITY

Augmented reality [7] is a technology that enables virtual images to be overlaid on views of the real world. Due to its capability to improve the perception of reality, to support teamwork, visual display of virtual objects, and to enable transitions between real and

virtual environments, the augmented reality can be used to create novel interfaces for face-to-face and remote collaboration. For example, AR enables users to see virtual representations of remote people in front of them and have spatial interactions with them, as if being there in person.

Wearable computers and cameras can be combined with augmented reality information display to support remote collaboration and to significantly improve performance on physical tasks. However, new research on interaction paradigms, presence and situational awareness needs to be conducted to create an augmented reality system that naturally enhances collaboration and establishes virtual co-location. In this case, situational awareness is defined as perception of given situation, its comprehension and prediction of its future state.

1.2. AFFECTIVE COMPUTING

Affective computing (AC) is a technology that "relates to, arises from, or deliberately influences emotions" [8]. Emotions guide cognition to enable adaptive responses to the environment, and can have a major impact on the perception, attention, memory and decision-making. Also, affect can have a significant impact on the driving behavior [9]. The affect reading by computer systems [1][2][10][3] is realized through the analysis of body gestures, facial expressions, speech and physiological signals.

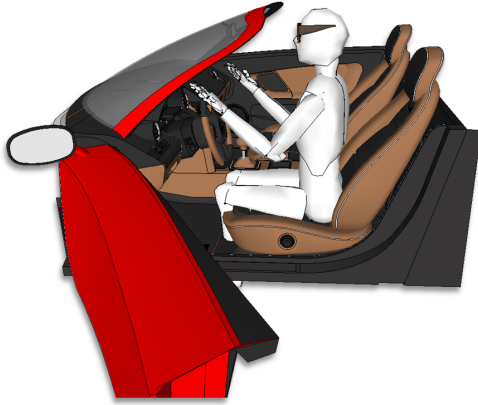


FIGURE 1. AR trainee using the car driving simulator

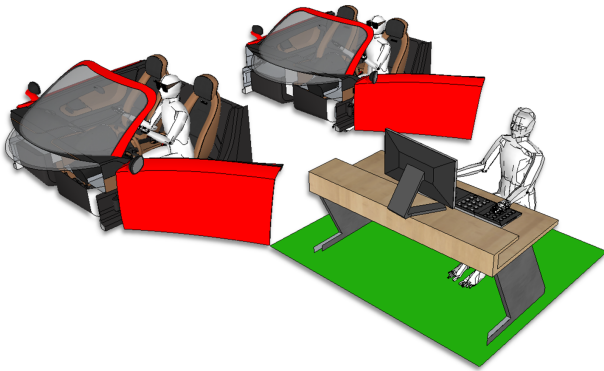


FIGURE 2. Training session including one trainer and two trainees.

1.3. HYBRID AR-AC FOR CAR DRIVING SIMULATIONS

The research aims to find solutions to specific challenges regarding integration of the affective computing technology with the augmented reality technology, with application in car driving simulations.

The motivation of this approach comes from lack of understanding on emotional awareness models in AR-based interactions between trainees and trainers, in a computer supported collaborative setup. Previous studies [4], [11] indicate advantages of combined AR/VR and AC for different scenarios.

The research aims to address the following questions:

- What are the best means for sensing and collecting affect data from the car drivers engaged in driving training sessions? Three sensing hardware will be investigated first, namely e-Health Sensor [12], Cortrium [13] and Empatica E4 wristband [14]. Multimodal approaches for affect recognition will be studied.
- Which models and (contact and non-contact) sensors are best suited for affect reading in car driving simulators?
- How can existent models be adapted in a data-fusion approach?, considering:

- ▷ Physiological sensors (i.e. heart rate, galvanic skin response, respiration rate, temperature, etc.),
- ▷ Facial expressions (from environmental cameras), with occlusion of the eye region by AR headset,
- ▷ Emotion in speech (prosody).

- How to build emotional awareness among trainees and the expert/the trainer? How to integrate and maintain emotional awareness?
- How to integrate and maintain emotional awareness and operation-specific situational awareness?
- What performance oriented model can be built to automatically adapt the AR system support (such as user interface, etc.) for a human-to-human interaction in virtual co-location, given the emotional awareness?
- In which way can (free-hands) HCI for AR/VR headsets benefit from emotional awareness, targeting the individual performance during collaboration?
- How to adapt existent interaction models?
- Which is the appropriate affect multimodal framework to support remote collaboration in AR/VR?
- How to adapt shared memory spaces to support affect-driven virtual environments?
- How to store, transfer and represent audio-video data and affect data, in a multi-user secure environment, with robustness to network breakdowns?
- What data-fusion models are to integrate multiple (marker-less) tracking (SLAM) independent systems for AR?
- What driving performance model can be built to automatically adapt the AR system support (such as the user interface, etc.) given the emotional awareness? Using this, for instance the AR-based simulation can be dynamically adapted according to the trainee's affect state, so that to increase drive learning performance.
- How to improve the (marker-less) tracking in the shared virtual environment to cope with rather high operational pace of security teams, and with large illumination variation?

- ▷ RGB-D tracking models

Section 2 presents a previous research on remote collaboration by augmented reality. Section 3 presents the model that integrates affective computing technology and an augmented reality technology. Section 4 details on AWARE, an affect-driven collaboration framework to support car driving simulations in augmented reality. Section 5 presents hardware and software components for affect reading during car driving simulation. The last section emphasizes on conclusions and future work for the research on combined augmented reality and affective computing technologies.

2. RELATED WORK

There have been car driving simulator-related approaches for rule-based systems that predicts in real-time the driver's intent [15] or for driver vigilance monitoring [16][17] applied to enhance safe driving experience.

Moeslund et al. [18] propose Arthur, an AR meeting system that permits multiple users wearing HMDs at a round table, to interact with objects specific to architecture and an urban planning domain. The interaction with augmented world is in two ways, using physical objects - placeholder objects and a wand, and by hand gestures.

Dong et al. [19] propose ARVita, an advanced collaborative AR tool with problem solving capabilities to be applied in classroom and in professional practice. In these scenarios, multiple users wearing HMDs and sitting around a table are able to perform interaction and to visualize dynamic simulations of engineering processes overlaid on the surface of the table.

The work of Wang and Dunston [17] advances two AR based systems for remote collaboration and a face-to-face co-located collaboration in the scenario of detecting design errors.

Jailly et al. [20] presents an AR system for enhancing the comprehension of the manipulated remote devices in distance learning domain that allow for communication between both students and teachers. Ferrise et al. [21] tackle the domain of maintenance operations of industrial product. In addition to using VR technology to support an operator to learn performing maintenance operations by combining traditional instruction manuals with simulation, the AR technology is employed to extend the scenario to tele-assistance. A VR-based skilled operator guides from the distance a trainee that is equipped with AR technology already displaying instructions on top of real product.

Nilsson et al. [22] propose an AR tool to improve collaboration between actors from different organizations such as the rescue services, the police and military personnel in a crisis management scenario while the same time sustaining individual needs.

Yabuki et al. [23] present a system in the early phase of development, aiming at supporting the cooperation between people working outdoor on environmental issues. The information provided to the users wearing HMDs relates to 3D representations of temperature distribution and wind distribution, velocity and direction.

Alem et al. [24] propose ReMoTe, a remote guiding system that integrates non-mediated hand gesture communication in the mining industry. The work scenario of the system implies the expert remotely assisting a worker using the hands to point to certain locations and to show specific manual procedures. Testing and validation of four early user interface design iterations aimed at maintenance tasks for repairing a photocopy machine, removing a card from

a computer mother board and assembling a Lego toy. Wichert [25] describes a mobile collaborative AR environment that uses web technologies. The collaborative environment allows a 3D game like Tetris to be played in real time by several users wearing HMDs. The players can be located in the same room, with possibility for extending the collaboration with a remote player. The game setup provides support for studying the two types of AR based collaboration: the co-located cooperative interaction with skilled workers, each having a different view of the AR world and the indirect interaction with remote expert that has the same view as the skilled worker.

In a similar way, Datcu et al. [4][5] propose an AR based scenario of playing a game collaboratively, to study complex problem solving by physically co-located and virtually co-located participants. Within the game, the goal of jointly building a tower of coloured blocks represents an approximation of a shared task. Individual expertise is modelled as possibility to move blocks of a distinct colour and shared expertise is modelled by possibility of all players to move blocks of same colour. By scaling down real-life, more complex problems, the study compares presence, workload and situational awareness in real world and AR collaboration scenarios. Additionally, Datcu et al. [6][7] developed a platform for tele-collaboration by AR for supporting teams in the security domain.

Schnier et al. [26] focus on studying the issues around establishing joint attention toward the same object or referent, in a physically co-located collaboration AR environment.

Gu et al. [27] conduct a study on the impact of 3D virtual representations and use of tangible user interfaces as support for synchronous design collaboration using the AR technology. The results indicate that the change from a physically co-located working environment to the virtual co-located scenario encourages the AR users to smoothly move between working on the same tasks and working on different tasks or different aspects of the design process.

The current state of the art on collaboration in AR provides relevant examples for AR based models that support a synchronous collaboration among users either being physically or virtually co-located, using free-hands or tangible interaction, static or mobile, or either using HMDs or other display devices. These research outcomes, however, have to be still investigated in car driving simulation domain, especially with regard to remote visualization, spatial interaction and remote authoring in training scenarios.

3. MODEL

Augmented reality systems are not limited to use of head-mounted devices and mainly have to combine real and virtual objects be interactive in real-time

and to register objects within the 3D. Due to the capability:

- to improve perception of reality,
- to support teamwork process,
- to support manual annotation by virtual objects,
- to support an interaction between virtual and augmented environments,

the technical solutions based on augmented reality have the potential to enhance novel interfaces for a computer-aided collaborative process, in face-to-face and remote collaboration scenarios.

Augmented reality systems can be used to establish the experience of being practically co-located by means of simulated presence. Augmented reality systems have been used to allow experts to spatially collaborate with others at any other place in the world without traveling and thereby creating the experience of being virtually co-located, e.g. in the field of a crime scene investigation [6],[28].

The affective computing technology makes use of measurements from (contact and non-contact) physiological body sensors to automatically recognize in real-time, the affect of the field personnel. The affect data is further on used to automatically:

- Improve interaction in augmented reality (for the field personnel),
- Increase immersion and situational awareness for spatially distributed users in the virtual reality [29] and augmented reality [30].

From the hardware point of view, a solution is to use a Microsoft HoloLens AR head mounted device (HMD) (see Figure 3) that has a depth sensor already integrated.



FIGURE 3. Standalone HoloLens AR device [31].

The investigation first considers a closed-loop AR-AC model for the virtual reality, proposed by Wu et al. [11]. Next, the closed-loop model is adapted for augmented reality.

3.1. CLOSED-LOOP ARCHITECTURE

According to the Yerkes-Dodson Law, the performance in mental tasks is dependent on arousal in the form of a non-monotonic function (Figure 4). The performance increases with arousal, given arousal is at low levels,

reaches the peak at a given arousal level and decreases after that optimal level.

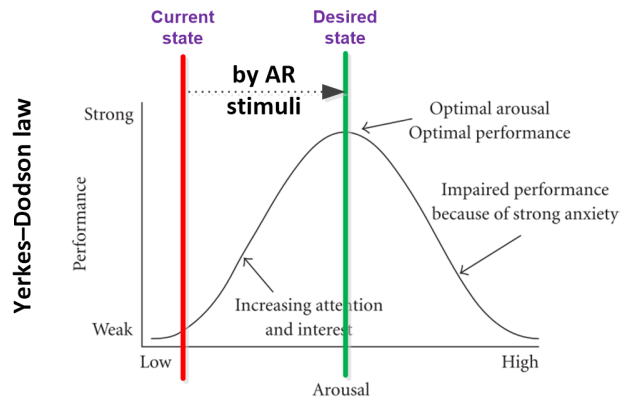


FIGURE 4. Yerkes-Dodson-Law

The closed-loop system consists of three components namely affect recognition component, affect-modeling component and affect control component. These components are displayed in Figure 5. The affective computer-aided collaboration approach based on augmented reality for car driving simulators, consists of three components namely:

- the affect recognition component,
- the affect-modeling component, and
- the affect control component.

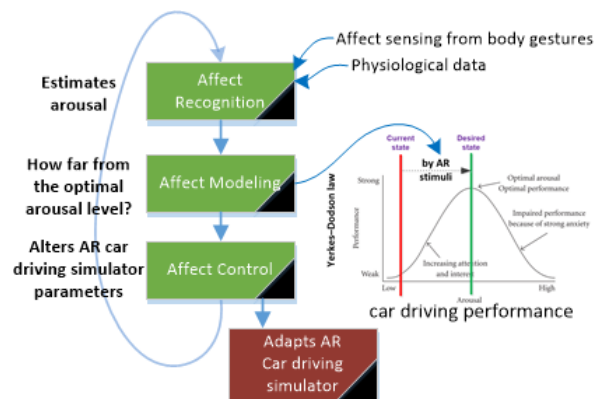


FIGURE 5. Closed-loop affective computing architecture for augmented reality.

3.1.1. AFFECT RECOGNITION

The affect recognition component performs the assessment of the user’s affect by analyzing different body signals such as psychophysiological signals and facial expressions, emotion in speech, etc. Ideally, these techniques to sense the trainee’s affect state should be working in real-time, should be automatic, non-contact, non-invasive and should have a high accuracy. Multi-modal approaches aggregate data from different types of sensors to decrease the error over estimated affect states.

3.1.2. AFFECT MODELING

The second component, the affect modeling creates a relationship between the trainee's affect and the features of the user's environment. This component determines how the trainee's affect should be changed, given her/his profile, the known arousal level for the optimal learning performance. This mapping is further used to identify car driving simulation parameters including the AR settings and the scenario and training session stimuli.

3.1.3. AFFECT CONTROL

The third component, the affect control provides the means for adapting the environment in such a way to get the trainee to the target affect state. This component can be semi-automatic, especially in scenarios including a human expert, such as in car driving simulation when the trainer can change the course of the training session with new stimuli that fit better the learning procedure. Given the car driving simulation parameters previously determined, this component applies selected changes. The aim is by applying these changes, the trainee's affect is changed towards the arousal state that is associated with optimal learning performance.

In order to optimize the car driving performance, the three components of sensing, modeling and control adjust the functionality of the car driving simulator according to the trainee's affect state.

4. AFFECTIVE FRAMEWORK FOR AUGMENTED REALITY

Virtual co-location relies on the augmented reality technology [32][33] to create spaces in which the trainer, the trainees and the objects are either virtually or physically present: it allows people to engage in spatial remote collaboration.

An affect-driven collaboration framework supports the virtual co-location technique for the trainee and for the trainer, the collection and processing (especially the data fusion) from the low/sensory level to the high semantic levels. In the following text, the trainee is referred to as the local user and the trainer is referred to as the remote expert. The framework is called Affective frameWork for Augmented REality - AWARE.

The AWARE framework has been developed with the goal to support computational demands and multi-modal data streams among running applications and data processing modules. AWARE is a highly scalable, modular and parallel environment for a distributed collaboration using AR (the diagram in Figure 6). It is developed to support virtual co-location of multiple users playing different roles in well-defined scenarios.

The following text describes AWARE in detail.

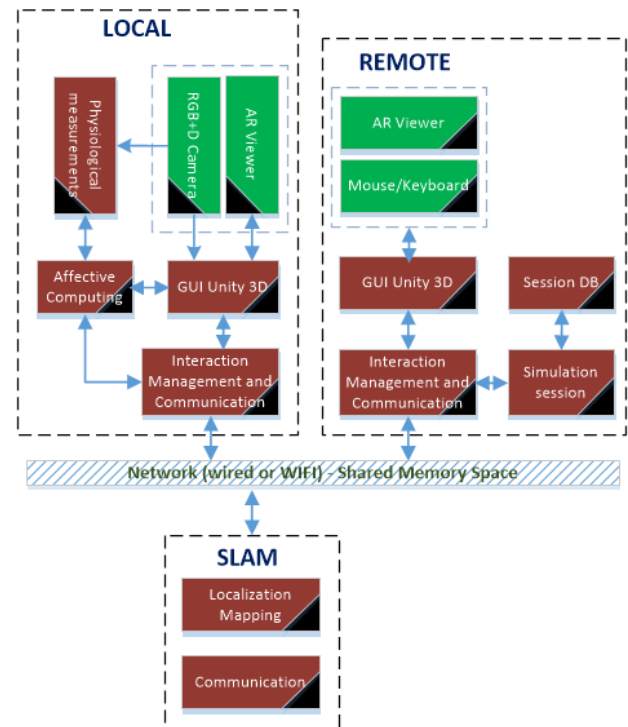


FIGURE 6. AWARE framework for adaptive AR training.

4.1. LOCAL AND REMOTE USER APPLICATIONS

AWARE is based on a centralized architecture. It contains different applications for local and remote users. The applications have different user interfaces which are created using the Unity game engine. The application for the local user is adapted for HMDs (standalone HoloLens optical see-through HMD in Figure 3).

The application for the remote user runs on a desktop computer or laptop. Thus, the remote user gets a screen-based visualization and can interact with the system by using the keyboard and a standard mouse device.

4.2. DIRECTIONS BY THE TRAINER

AWARE supports the collaboration process by providing the trainer (remote user) with tools to give directions in form of spatial annotations. The annotations appear in the view of the trainee (local user). The remote user can augment the view of the local user with the following elements:

- 3D-aligned objects (arrow/cube/sphere),
- 2D-screen aligned content (text, photo and video),
- dialog boxes for introducing text for the graphical objects and also for colours of some visual elements,
- screen aligned counter and text,
- screen aligned (flashing) panic button in the form of a text with the frame,
- screen aligned image and colour border (to show info about a person), and

- virtual stickers (indicating scanning area at the crime scene).

All annotations are encoded as distinct data messages and events that pass through shared memory system.

Updates are automatically sent to each software module or application (such as the software application of the local user or the software application of the remote user) by using a notification and a push system of events and data.

4.3. CONSISTENCY OF ACTIONS

The consistency of actions is a critical aspect for establishing a virtual co-location. For that purpose, local updates are processed only when data and events are received as a feedback from the AWARE shared memory space module. Consequently, the application for the trainer (remote user) does not execute user input updates in the graphical user interface immediately. Such updates are applied only after the data and events of the user input are available in the shared memory space. The flow of data and event notifications are illustrated in the diagram in Figure 7.

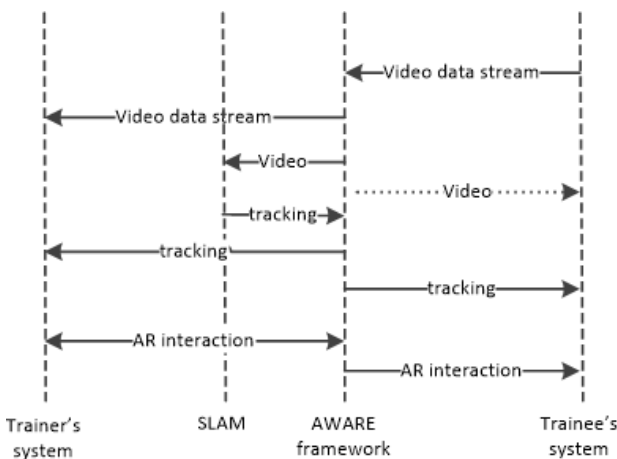


FIGURE 7. Diagram of data and event notifications for user actions between AWARE modules and user applications.

4.4. DATA AND EVENT NOTIFICATIONS THROUGH AWARE

The data communication is established via the shared memory mechanism of AWARE. The shared memory mechanism handles parallel connections from different users located in different physical environments, and allows data sharing across different types of devices, including mobile AR HMD systems.

In AWARE, the network communication is implemented using both TCP/IP standard - for the data transfers between the server and software modules running on hardware devices linked via network cables, and UDP standard - for data transfers between software modules connected via wireless links. In case

of the UDP-based network communication, each frame from video sequence is encoded as a compressed image (using VGA resolution and JPEG encoding) into a UDP packet.

4.5. SHARED VISUALIZATION

Virtual co-location is enabled by sharing a video channel from the HMD of the trainee (local user) with the trainer (remote user). The shared video stream, thereby, forms basis for the shared work environment. The shared memory mechanism provides functionality for synchronizing the video frames with the updates from the HMD camera tracking and with the manual annotations made by the remote user.

Via a graphical user interface, the remote user can modify the AR content, which is visible on the screen. The AR system has also capability to record, store and display a photo and video content. The AWARE framework provides functionality to store data remotely, not only locally. Such data are persistent as long as the AWARE server is up and running.

Especially for videos, this mechanism ensures synchronization of the display: at a time moment, the same video frame is played for both the local and the remote user. The photos and videos are stored in separate files using standard formats, such as .png, .jpg and .avi.

The transmission of video streams makes use of the Motion JPEG - MJPEG video compression format. Each frame of the video sequence is compressed and transferred as a JPEG image.

The automatic process of event and data notification takes place only for the software modules and applications that register for the specific type of data. In this way, for instance the application for the local user does not have to receive, via the network, video from its own camera (it has a direct access to this video).

4.6. COMMUNICATION DECOUPLING

AWARE achieves decoupling of communication by supporting both local user's and remote user's applications to access updates through a shared memory space. This functionality is tightly coupled to the virtual co-location paradigm that enables multiple users at different physical locations to work collaboratively according to their roles in a well-defined scenario. Collaboration at a distance is possible as long as both users have a network connectivity to the AWARE server.

4.7. LOCAL USER TRACKING IN THE PHYSICAL ENVIRONMENT

The local user's motion in the physical environment is automatically tracked in an augmented reality by using a SLAM system. A SLAM system (Simultaneous Localization And Mapping) generates and updates a map of an unknown physical environment while simultaneously keeping track of a user's location within it.

AWARE can run different SLAM systems. One of them is RDSLAM [34], a real-time marker-less monovision SLAM more suited for indoor environments developed at Zhejiang University in Hangzhou, China. Another SLAM is LSD-SLAM [35], a real-time marker-less SLAM running on semi-dense depth maps. A third SLAM system is ORB-SLAM2 [36], a real-time system for monocular, stereo and RGB-D cameras that computes camera trajectory and a sparse 3D reconstruction.

The tracking module is integrated as a module of the AWARE framework and can run on computer of the trainer (local user) or on a separate hardware system.

In addition to tracking the HMD position and orientation, SLAM performs a mapping of the physical environment (of the trainer) and generates an internal 3D representation of the physical world.

The physical world is represented in terms of data structures related to key points discovered during the tracking, and to recordings of a camera position and orientation with respect to the key-points. Such data are mainly used internally by the SLAM during the tracking process.

The tracking procedure identifies a set of key points (as natural visual features) from each frame of the video sequence. Estimation of the camera parameters (location and orientation) keeps track of all the key points during the time. This means newly discovered key points are stored in the system memory and already-known key-points are tracked when detected. As a part of the AWARE functionalities, the tracking module stores all detected key-points in the current frame of the digital video sequence, the camera location and orientation data in the shared memory space. Once these data are stored on the shared memory space, all the system modules are notified on the updates so that the data can be further read, if necessary. This mechanism is illustrated in Figure 7.

The result of a markerless tracking provides the HMD camera location and orientation while the mapping result provides a representation of the physical world in form of a sparse cloud of 3D points.

4.8. SELECTION AND POSITIONING OF 3D VIRTUAL OBJECTS

The sparse cloud of 3D points represents visual key-points, which connect the augmented world to the physical world and further act as virtual anchors supporting annotation by AR markers.

The remote user can attach a virtual 3D object by using the user interface. After choosing a desired type of augmentation (by selecting a specific icon) from the bar of icons on the top-left side of the user interface, the remote user proceeds by directly mouse-clicking on the screen showing the current frame of the video sequence.

The mouse click event associates the 2D coordinates of the mouse cursor on the laptop screen. In order to attach the selected 3D virtual object, the 2D mouse cursor location has to be converted to 3D coordinates, related to the reference system of the SLAM tracking. This conversion is done by requesting the SLAM module via the shared memory space implemented in AWARE, to map the 2D coordinate in the current video frame to the 3D coordinate. The mapping implies searching for the closest key point from the equivalent 3D location of the 2D point of the mouse click event generated through the user interface of the remote expert. Once computed, the closest 3D point in the coordinate reference system of SLAM is sent back to the expert system application via the shared memory space. This is illustrated in the diagram in Figure 8.

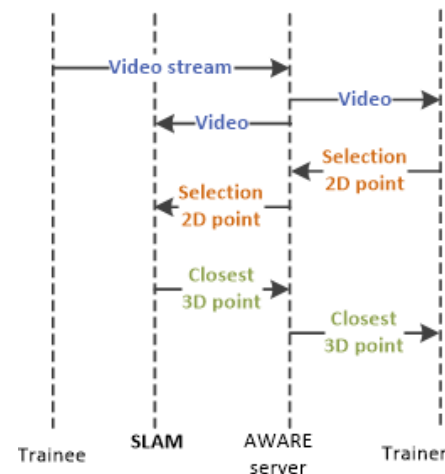


FIGURE 8. Diagram of data and event notification for user actions within AWARE modules and applications.

Once attached to key point, a 3D-aligned virtual object is correctly rendered in the next video frames on the user interfaces for both local and remote conditions during run-time, the generation of graphical content being consistent with the HMD camera motion (Figure 4, the upper part of the diagram).

5. SETUP

5.1. HARDWARE

The affect adaptive AWARE framework for training supports hardware equipment for the augmented reality and physiological sensors. Such equipment includes Microsoft Hololens headset (Figure 3), and physiological sensors such as: physiological eHealth kit (Figure 9), heart rate detector Cortrium (Figure 10) and Empatica E4 (Figure 11).

The e-Health Sensor Shield V2.0 [12] (Figure 9) can perform a real-time body monitoring for biometric and medical applications by using 10 different sensors including: pulse, oxygen in blood (SPO2), airflow (breathing), body temperature, elec-

trocardiogram (ECG), glucometer, galvanic skin response (GSR - sweating), blood pressure (sphygmomanometer), patient position (accelerometer) and muscle/electromyography sensor (EMG).



FIGURE 9. eHealth platform [12].

Cortrium [13] (Figure 10) can assess body surface temperature, activity, and respiration rate and it also contains a high-performance three-channel ECG for screening and diagnostics of cardiological diseases.

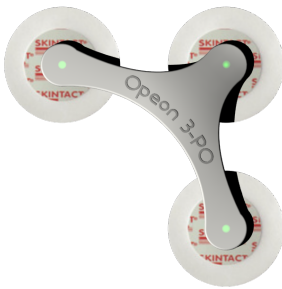


FIGURE 10. Cortrium heart rate monitor [13].

Empatica E4 [14] (Figure 11) is a real-time wristband monitor of the physiological signals. The device incorporates a photoplethysmography sensor to measure blood volume pulse (providing estimations on the heart rate, heart rate variability and other cardiovascular features). In addition, it has a 3-axis accelerometer, and an infrared thermopile that reads peripheral skin temperature. An built-in electrodermal activity sensor measures the sympathetic nervous system arousal and derives features related to stress, engagement, and excitement.

5.2. AFFECT RECOGNITION

5.2.1. NON-CONTACT HEART RATE DETECTION

Non-contact analysis of physiological indicators has a major impact on several application domains making use of live monitoring of the people's faces, in realistic scenarios involving working, playing and resting. Photoplethysmography represents a noncontact, non-invasive and low-cost method that makes use of variations of transmitted or reflected light to determine



FIGURE 11. Empatica E4 wristband physiological signal monitoring [14].

cardiovascular blood volume. Considering normal ambient light as illumination source, allows for using video cameras and even regular webcams as heart rate detection sensors.

According to the previous study [37], specific face skin areas provide enough information to estimate the heart rate, by using computer vision techniques only, including the Viola&Jones object detector and Active Appearance Models. The results of such a non-contact face analysis method for the heart rate detection are depicted in Figure 12. The tests have proved that the whole face area does not necessarily provide optimal region of interest for the pulse detection, the best being the cheek and forehead regions.

The findings can be used to optimize the scanning of the face region which in turn leads to shorter computation times and more accurate results for the heart rate detection. In addition, lower face features can be used to sample color information for the analysis while the trainee in the car driving simulator wears an augmented reality HMD such as HoloLens.

5.2.2. BI-MODAL EMOTION RECOGNITION

The assessment of emotional levels from speech can be naturally done by identifying patterns in the audio data and by using them in a classification setup. The features we extract are the energy component and 12 mel-frequency cepstral coefficients together with their delta and the acceleration terms. HMMs models makes use of Gaussian mixtures with different number of components. The evaluation indicates the most efficient HMM model makes use of 4 states and 40 Gaussians per mixture. The accuracy of this HMM-based classifier is 55.90% [38] (Table 1).

Making facial expression recognizers with hidden Markov models and Local Binary Pattern features (LBPs) implies the identification of the optimal model parameters. Finding the best number of states, the best number of Gaussian mixture components and the best set of local binary patterns, is not a trivial task. We start from the results of the Adaboost.M2 classifiers. For evaluation of the facial expression recognition, we have generated HMM models for each

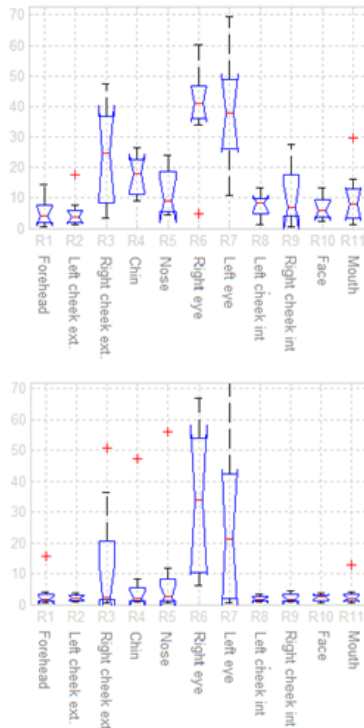


FIGURE 12. Box plots of heart rate median errors together with the 25th and 75th percentiles, for each face region.

emotion category. The best facial expression recognition model uses 268 distinct features that correspond to a selection of 45 features from each facial expression category. The accuracy of this classifier is 37.71% [38] (Table 2).

The recognition of emotions based on a decision level fusion implies the combination of final classification results obtained by each modality separately. For this, we take into consideration four sets of unimodal classification results namely from the speech-oriented analysis and from the separate LBP-oriented analysis which use visual features from the whole face image from specific face regions. We use these sets together with a weight function that allows for setting different importance levels for each set of unimodal results. This weight-based semantic fusion models asynchronously the emotion in visual and auditory channels. The best model obtained in this way has the accuracy of 56.27% [38].

6. CONCLUSIONS

This is the first research to study a role of the affect in an augmented reality for the car driving simulation. The augmented reality technology for a remote collaboration by virtual co-location, is used to support the trainer and the trainee during the car driving training session. In order to improve learning performance during training, the affective computing technology is used to sense the trainee's affect state and to further update semi-automatically car driving simulator

parameters. The affect recognition consists of a multimodal technique that processes inputs from contact-based physiological sensors, from facial expressions readings and from the emotion in speech. The way of using the affective computing technology to further control the car driving simulation in augmented reality, stands for a novel research topic worldwide.

Future work will focus on preparing a fully functional car simulation prototype and on conducting a series of experiments. The research findings will contribute to a deeper knowledge on integration of situational and emotional awareness in the augmented reality for the car driving training simulation and training.

REFERENCES

- [1] D. Datcu, L. J. M. Rothkrantz. Emotion recognition using bimodal data fusion. In *Proceedings of the 12th International Conference on Computer Systems and Technologies*, CompSysTech '11, pp. 122–128. ACM, New York, NY, USA, 2011. DOI:10.1145/2023607.2023629.
- [2] L. Rothkrantz, D. Datcu, I. Chiriacescu, A. Chitu. Assessment of the emotional states of students during e-learning. In *Proceedings of the International Conference on e-Learning and the Knowledge Society*, e-Learning '09, pp. 77–82. 2009.
- [3] D. Datcu, L. J. M. Rothkrantz. Automatic bi-modal emotion recognition system based on fusion of facial expressions and emotion extraction from speech. In *Proceedings of the 8th IEEE International Conference on Automatic Face & Gesture Recognition*, FG '08. IEEE, 2008. DOI:10.1109/AFGR.2008.4813335.
- [4] D. Datcu. On the enhancement of augmented reality-based tele-collaboration with affective computing technology. In *Proceedings of the 11th Romanian Human-Computer Interaction Conference, Romania*, RoCHI'14. 2014.
- [5] D. Datcu, M. Cidota, S. Lukosch, et al. Virtual co-location to support remote assistance for inflight maintenance in ground training for space missions. In *Proceedings of the 15th International Conference on Computer Systems and Technologies*, CompSysTech '14, pp. 134–141. ACM, New York, NY, USA, 2014. DOI:10.1145/2659532.2659647.
- [6] S. Lukosch, H. Lukosch, D. Datcu, M. Cidota. Providing information on the spot: Using augmented reality for situational awareness in the security domain. *Computer Supported Cooperative Work (CSCW)* **24**(6):613–664, 2015. DOI:10.1007/s10606-015-9235-4.
- [7] R. T. Azuma. A survey of augmented reality. *Presence: Teleoper Virtual Environ* **6**(4):355–385, 1997. DOI:10.1162/pres.1997.6.4.355.
- [8] R. W. Picard. *Affective Computing*. MIT Press, Cambridge, MA, USA, 1997.
- [9] T.-Y. Hu, X. Xie, J. Li. Negative or positive? the effect of emotion and mood on risky driving. *Transportation Research Part F: Traffic Psychology and Behaviour* **16**:29 – 40, 2013. DOI:http://dx.doi.org/10.1016/j.trf.2012.08.009.
- [10] L. J. Rothkrantz, D. Datcu. Assessment of emotion states during e-learning. *Journal of Communication and Cognition* **43**(1-2), 2010.

(%)	Fear	Surprise	Sadness	Anger	Disgust	Happy
Fear	91.72	2.07	0.69	1.38	2.76	1.38
Surprise	24.11	44.68	9.22	11.35	4.26	6.38
Sadness	25.19	14.81	41.48	6.67	5.19	6.67
Anger	19.38	18.60	3.88	48.06	6.98	3.10
Disgust	23.78	9.09	8.39	8.39	38.46	11.89
Happy	5.84	5.84	10.22	2.19	6.57	69.34

TABLE 1. The confusion matrix of the HMM that has 4 states and 40 Gaussian components; the accuracy of the emotion recognition from speech model is 55.90% for six basic emotion categories.

(%)	Anger	Disgust	Happy	Surprise	Sadness	Fear
Anger	18.62	15.86	11.03	24.13	17.24	13.10
Disgust	9.92	60.28	10.63	63.82	6.38	6.38
Happy	6.20	17.05	48.06	18.60	5.42	4.65
Surprise	10.48	2.79	9.09	53.14	16.08	8.39
Sadness	17.51	10.94	10.94	25.54	19.70	15.32
Fear	9.62	16.29	6.66	26.66	14.07	26.66

TABLE 2. Confusion matrix of the best HMM facial expression classifier using LBP features.

- [11] D. Wu, C. G. Courtney, B. J. Lance, et al. Optimal arousal identification and classification for affective computing using physiological signals: Virtual reality stroop task. *IEEE Transactions on Affective Computing* 1(2):109–118, 2010. DOI:10.1109/T-AFFC.2010.12.
- [12] e-health sensor platform v2.0 for arduino and raspberry pi, <https://www.cooking-hacks.com/documentation/tutorials/ehealth-biometric-sensor-platform-arduino-raspberry-pi-medical>.
- [13] Cortrium heart rate monitor, <http://cortrium.com/>.
- [14] Empatica e4 wristband, <https://www.empatica.com/e4-wristband>.
- [15] M.-I. Toma, D. Datcu. *Determining Car Driver Interaction Intent through Analysis of Behavior Patterns*, pp. 113–120. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012. DOI:10.1007/978-3-642-28255-3_13.
- [16] H. J. Dijkers, M. A. Spaans, D. Datcu, et al. Facial recognition system for driver vigilance monitoring. In *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, vol. 4, pp. 3787–3792 vol.4. 2004. DOI:10.1109/ICSMC.2004.1400934.
- [17] D. Datcu, L. J. M. Rothkrantz. Multimodal recognition of emotions in car environments. In *Driver Car Interaction & Interface 2009*. 2009.
- [18] T. B. Moeslund, M. Störring, W. Broll, et al. The arthur system: An augmented round table. *Journal of Virtual Reality and Broadcasting* 34:2003.
- [19] S. Dong, A. H. Behzadan, F. Chen, V. R. Kamat. Collaborative visualization of engineering processes using tabletop augmented reality. *Advances in Engineering Software* 55:45 – 55, 2013. DOI:<http://dx.doi.org/10.1016/j.advengsoft.2012.09.001>.
- [20] X. Wang, P. S. Dunston. Comparative effectiveness of mixed reality based virtual environments in collaborative design. In *Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on*, pp. 3569–3574. 2009. DOI:10.1109/ICSMC.2009.5346691.
- [21] F. Ferrise, G. Caruso, M. Bordegoni. Multimodal training and tele-assistance systems for the maintenance of industrial products. *Virtual and Physical Prototyping* 8(2):113–126, 2013. <http://dx.doi.org/10.1080/17452759.2013.798764> DOI:10.1080/17452759.2013.798764.
- [22] S. Nilsson, B. Johansson, A. Jonsson. Using ar to support cross-organisational collaboration in dynamic tasks. In *Mixed and Augmented Reality, 2009. ISMAR 2009. 8th IEEE International Symposium on*, pp. 3–12. 2009. DOI:10.1109/ISMAR.2009.5336522.
- [23] N. Yabuki, S. Furubayashi, Y. Hamada, T. Fukuda. *Collaborative Visualization of Environmental Simulation Result and Sensing Data Using Augmented Reality*, pp. 227–230. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012. DOI:10.1007/978-3-642-32609-7_32.
- [24] L. Alem, F. Tecchia, W. Huang. Remote Tele-assistance System for Maintenance Operators in Mines. 2011.
- [25] R. Wichert. Collaborative gaming in a mobile augmented reality environment. In *Proceedings of EUROGRAPHICS - Ibero-American Symposium in Computer Graphics - SIACG 2002*, pp. 31–37. 2002.
- [26] C. Schnier, K. Pitsch, A. Dierker, T. Hermann. *Collaboration in Augmented Reality: How to establish coordination and joint attention?*, pp. 405–416. Springer London, London, 2011. DOI:10.1007/978-0-85729-913-0_22.
- [27] N. Gu, M. J. Kim, M. L. Maher. Technological advancements in synchronous collaboration: The effect of 3d virtual worlds and tangible user interfaces on architectural design. *Automation in Construction* 20(3):270 – 278, 2011. Augmented and Virtual Reality in Architecture, Engineering and Construction (CONVR2009), DOI:<http://dx.doi.org/10.1016/j.autcon.2010.10.004>.

- [28] D. Datcu, S. Lukosch, H. Lukosch, M. Cidota. Using augmented reality for supporting information exchange in teams from the security domain. *Security Informatics* **4**(1):1–17, 2015. doi:10.1186/s13388-015-0025-9.
- [29] T. Gross. Supporting effortless coordination: 25 years of awareness research. *Computer Supported Cooperative Work (CSCW)* **22**(4):425–474, 2013. doi:10.1007/s10606-013-9190-x.
- [30] S. S. Ayyagari, K. Gupta, M. Tait, M. Billinghamurst. Cosense: Creating shared emotional experiences. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '15, pp. 2007–2012. ACM, New York, NY, USA, 2015. doi:10.1145/2702613.2732839.
- [31] Microsoft hololens, <https://www.microsoft.com/microsoft-hololens/en-us>.
- [32] R. Azuma, Y. Baillot, R. Behringer, et al. Recent advances in augmented reality. *IEEE Computer Graphics and Applications* **21**(6):34–47, 2001. doi:10.1109/38.963459.
- [33] R. T. Azuma. A survey of augmented reality. *Presence: Teleoper Virtual Environ* **6**(4):355–385, 1997. doi:10.1162/pres.1997.6.4.355.
- [34] W. Tan, H. Liu, Z. Dong, et al. Robust monocular slam in dynamic environments. In *Mixed and Augmented Reality (ISMAR), 2013 IEEE International Symposium on*, pp. 209–218. 2013. doi:10.1109/ISMAR.2013.6671781.
- [35] J. Engel, T. Schöps, D. Cremers. *LSD-SLAM: Large-Scale Direct Monocular SLAM*, pp. 834–849. Springer International Publishing, Cham, 2014. doi:10.1007/978-3-319-10605-2_54.
- [36] R. Mur-Artal, J. M. M. Montiel, J. D. Tardós. ORB-SLAM: a versatile and accurate monocular SLAM system. *IEEE Transactions on Robotics* **31**(5):1147–1163, 2015. doi:10.1109/TRO.2015.2463671.
- [37] D. Datcu, M. Cidota, S. Lukosch, L. Rothkrantz. Noncontact automatic heart rate analysis in visible spectrum by specific face regions. In *Proceedings of the 14th International Conference on Computer Systems and Technologies*, CompSysTech '13, pp. 120–127. ACM, New York, NY, USA, 2013. doi:10.1145/2516775.2516805.
- [38] D. Datcu. *Multimodal recognition of emotions*. PhD. Thesis. Delft. Delft University of Technology, 2009.