Deliverable

WP3 – Development of short-range 3D-imaging systems

D3.14 Demonstrator AS2 of a wide-angle in-cabin automotive system

Project Information

Grant Agreement n°	826600
Dates	01/05/19 - 31/10/22

This project has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 826600. The JU receives support from the European Union's Horizon 2020 research and innovation programme and France, Sweden, Greece, Spain, United Kingdom, Germany, Luxembourg, Latvia, Hungary.
The VIZTA project results presented reflect only the author's view. The JU is not responsible for any use that may be made of the information it contains.

Document status

Document Information

Deliverable name	VIZTA_D3.14_VF
Responsible beneficiary	IEE S.A.
Contributing beneficiaries	DFKI and IEE S.A.
Contractual delivery date	M42 – 30/10/2022
Actual delivery date	M42 – 26/10/2022
Dissemination level	Public

Document approval

Name	Position in project	Organization	Date	Visa
Laurent Dugoujon	WP3 Leader	ST SAS C2	25/10/2022	OK
Laurent Dugoujon	Coordinator	ST SAS C2	25/10/2022	OK
Fabienne Brutin	PMO member	Benkei	26/10/2022	OK

Document history

Version	Date	Modifications	Authors / Organization
VF	25/10/2022	Final version, reviewed OK by L.Dugoujon	Y. Anisimov, J. Katrolia, B. Mirbach, J. Rambach / DFKI, F. Grandidier / IEE

Table of content

Executive summary

1 Description of the deliverable objective and content

The deliverable describes the demonstrator AS2 of a wide-angle in-cabin automotive system developed within VIZTA in close collaboration by [IEE] and [DFKI]. First, this report gives a short overview over the demonstrator set-up at [DFKI], the hard- and software architecture, as well as the use case, camera hardware, and processing platform. Then, the deliverable report summarizes the results of the hardware and algorithm development for this demonstrator which were already documented in previous deliverables D3.12 and D3.13 as well as the results achieved in the final task T.C.6, for completing the demonstrator. This task consisted in integrating the algorithm and software components in a demonstrator software pipeline, to deploy it on the demonstrator processing platform, and moreover, to adapt the algorithm to the modified S2 evaluation kit.

The main achievement in the demonstrator hardware development was the modification of the S2 evaluation kit delivered by [ST GNB2 SAS] (see D3.6) to meet the requirements of the application in terms of required field of view (see D3.32). This modification was performed by [IEE] in close collaboration with [DFKI] and [ST GNB2 SAS]. All details and results quantifying the image quality of the final hardware are reported in D3.13. **Erreur ! Source du renvoi introuvable.** shows design and picture of this camera together with a picture of the Kinect Azure and a point cloud recorded with it.

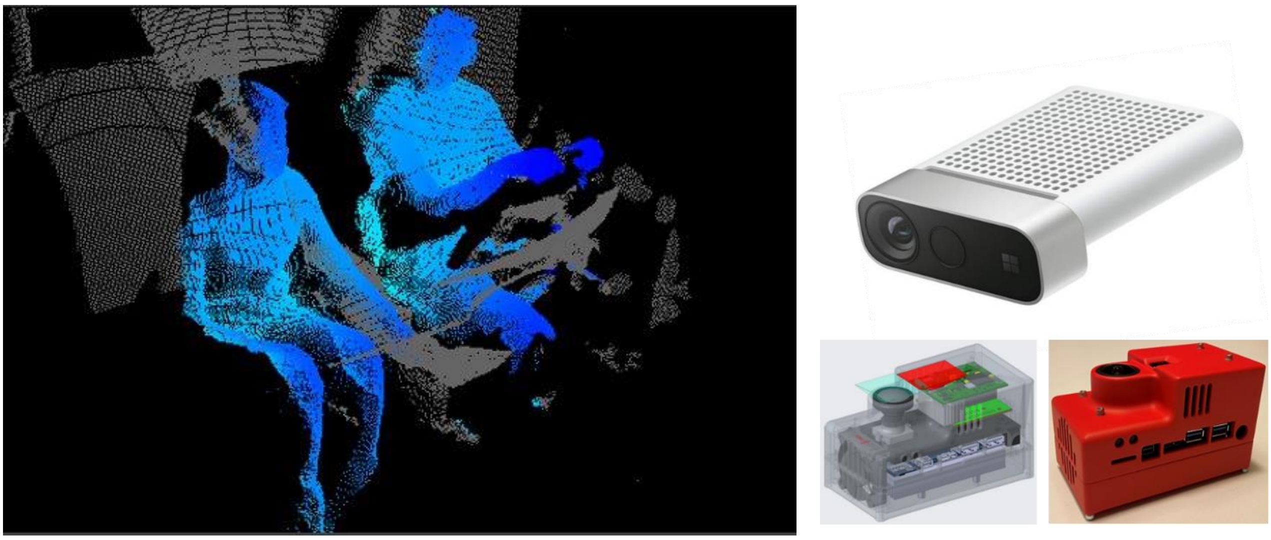

Figure 1: 3D iToF cameras used to realize the in-cabin demonstrator AS2: Right: Top: Kinect Azure, Bottom: Modified S2 camera developed within VIZTA (design and prototype). Left: in-cabin point cloud recorded with KInect Azure

The concept of the deep-learning algorithm performing the core in-cabin monitoring tasks, namely person and object detection and segmentation based on depth images grabbed by a wide-field-of view ToF-camera, had already been presented in D3.34. The algorithm selected by [DFKI] after implementing, adapting, and evaluating several state-of-the art approaches on the benchmark dataset generated within VIZTA, had then been optimized and deployed on the selected target platform for real-time demonstration. Delivery D3.12 describes this algorithm, including its architecture, the optimization and key performance figures. This algorithm has finally been to the modified S2 sensor which comprised an adaptation of the pre-processing as well as the recording of training and test data with the S2 sensor for retraining the network model of the algorithm

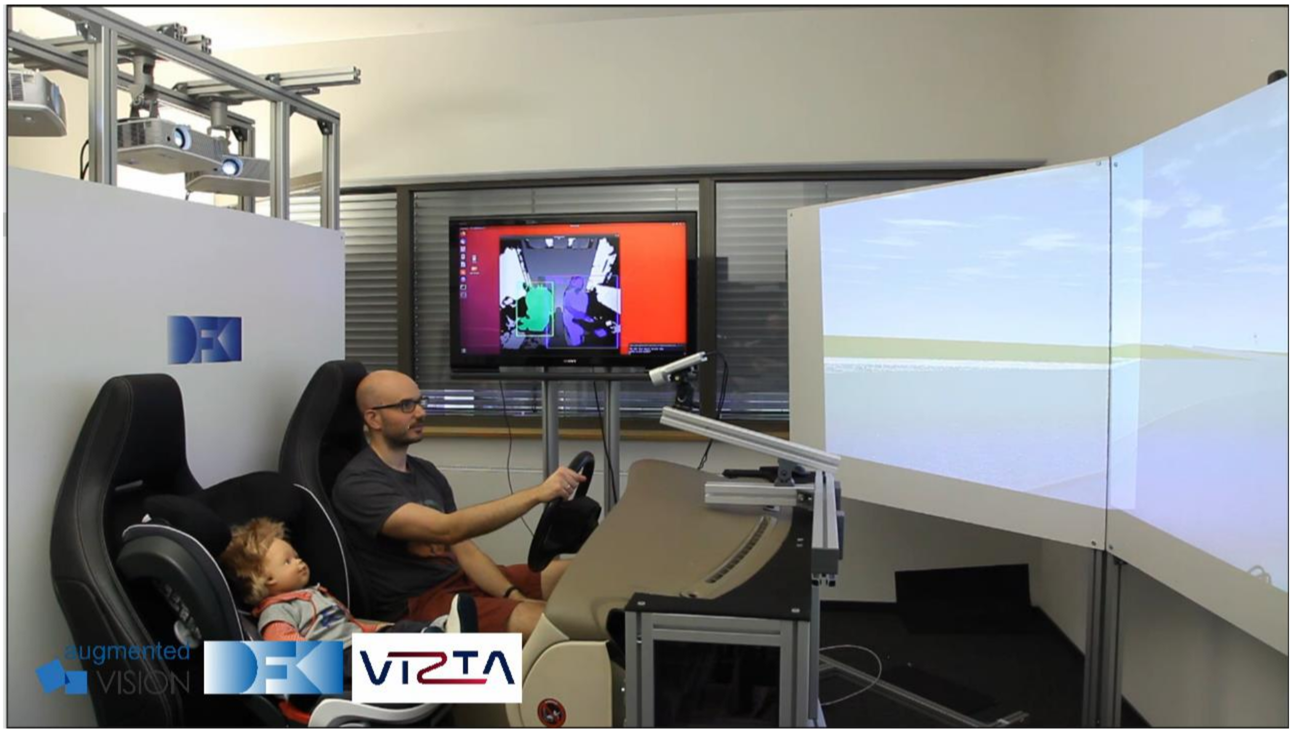

Figure 2: The final demonstrator in action.

Figure 2**Erreur ! Source du renvoi introuvable.** shows the final demonstrator in action. The graphical user interface is displaying the algorithm output. It shows the undistorted depth image as grayscale image together with the detection, classification, and segmentation result of the algorithm.

2 Brief description of the state of the art

Dedicated in-cabin test set-ups described in the literature often provide only 2D-videos from confined areas in the vehicle, where either the driver's head or hand is located or 3D annotations are missing (see, e.g. the VIVA challenge[1]). Meanwhile driver monitoring systems (DMS), being 2D-cameras pointing on the drivers face and providing information on fatigue or distraction have been introduced into vehicle to support advanced driver assistant systems (ADAS)[2].

[1] VIVA: Vision for intelligent vehicles and applications, http://cvrr.ucsd.edu/vivachallenge/ (2016), last accessed 2019/07/25

2 https://www.telematicswire.net/the-adas-and-dms-explainer/

The development of time-of-flight based in cabin monitoring is relatively new. Most recently several in-cabin set-ups with time-of-flight cameras and benchmark datasets were presented targeting the development of various functions as driver head pose estimation[34], driver action recognition[5] or driver anomaly detection[6]. These setups use only time of flight with narrow field of view, not able to monitor the entire front seats of the vehicle cabin to realize an occupant classification and body pose estimation of both driver and passenger. The same limitation applies for already available 3D-camera-based user interface systems, introduced by BMW (2017)[7] and Daimler (2020)[8], These systems are integrated in the overhead module of the car and can recognize hand, respectively arm gestures, in limited central areas of the vehicle cabin. To be able to monitor the full in-cabin scene, automotive supplier LG[9] has recently proposed to combine the ToF camera with a wide field of view RGB and IR camera.

With the demonstrator AS2 we avoid such a complex solution by showing that it is possible to monitor from the preferred position in the overhead module the entire in-cabin scenario on the front seats with a single wide-field of view ToF camera with a resolution larger than VGA.

On algorithm side, deep neural network-based approaches have in recent years become the state of the art in object detection and are a highly dynamic field of research. In D3.34 we have already presented an exhaustive overview of the latest state of the art methods and assessed several of them for our purpose. The selected approach is based on the YOLACT architecture, first published in 2019[10]. This approach proved to be more accurate while being less demanding in run time and memory needs. Nevertheless, to deploy a neural network like YOLACT on embedded platform dedicated for AI application, the algorithm had to be optimized, e.g., by quantization and pruning of network weights. This state-of-the art optimizer for NVIDIA platforms is TensorRT[11] , an SDK for high-performance deep learning inference, which has been also used for the
optimization of the algorithm, is described herein.

---

[3] M. Roth and D. M. Gavrila. DD-pose - a large-scale driver head pose benchmark, In IEEE Intelligent Vehicles Symposium, 2019.

[4] A. Schwarz, M. Haurilet, M. Martinez, and R. Stiefelhagen. Driveahead - a largescale driver head pose dataset. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[5] M. Martin *et al*., "Drive&Act: A Multi-Modal Dataset for Fine-Grained Driver Behavior Recognition in Autonomous Vehicles," *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 2801-2810, doi: 10.1109/ICCV.2019.00289.

[6] O. Köpüklü, J. Zheng, H. Xu and G. Rigoll, "Driver Anomaly Detection: A Dataset and Contrastive Learning Approach," *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2021, pp. 91-100, doi: 10.1109/WACV48630.2021.00014.

[7] https://www.youtube.com/watch?v=rT6RmINmXR0

[8] https://group-media.mercedes-benz.com/marsMediaSite/en/instance/ko.xhtml?oid=46757031

[9] https://www.lg.com/global/mobility/cabin-camera, retrieved 26/09/2022

[10] D. Bolya, C. Zhou, F. Xiao and Y. J. Lee, "YOLACT: Real-Time Instance Segmentation," *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 9156-9165, doi: 10.1109/ICCV.2019.00925.

[11] https://developer.nvidia.com/tensorrt

3 Deviation from objectives and corrective actions

Due to software integration problems of the S2 sensor into the demonstrator pipeline at [DFKI] the demonstrator is not running in online mode with the integrated S2 sensor.
As contingency measure, there is the online demonstrator AS2 with the consumer camera Kinect Azure plus an offline version of it allowing to process S2 data. The adaptation of the algorithm to the S2 data has, however, been accomplished with very good results.

4 Impact of the results

The developed demonstrator AS2 is the first wide angle time-of-flight based system able to monitor the entire front cabin of a vehicle. This demonstrator illustrates the advantages of the time-of-flight camera technology for this use case, being the accurate detection and classification of vehicle occupants based on their 3D contour.

The algorithms developed within VIZTA by [DFKI] and implemented in the demonstrator are among the first deep learning approaches that use exclusively or predominantly depth information. The scientific publications resulting from the development on the algorithm approaches[12],[13] as well as the VIZTA TICaM dataset[14] benchmark dataset, are expected to foster research and development of novel deep learning algorithms for in-cabin monitoring functions based on ToF camera systems. An important feature of the algorithm is also the easy scalability, e.g., to more different object classes, and applicable to different cameras. With this deliverable we have moreover shown that the developed high-performance algorithm can be deployed on a cost and energy optimized automotive qualified processing platforms dedicated for AI applications[15], as they are becoming state of the art in future autonomous vehicles.

The development of this final demonstrator hardware has moreover exemplified that the S2 sensor can be integrated in an application specific ToF camera by customizing it with off-the-shelf optical components, and thereby fulfil high requirements towards resolution, noise, and framerate. In this way the final hardware and the demonstrator AS2 shall foster in the coming years the development of ToF camera based in-cabin sensing solutions for safety, comfort and human-vehicle interfaces that go beyond existing hand/arm-gesture recognition. A sensor fusion of the ToF-based demonstrator with a driver monitoring system (DMS) is in addition a clearly indicated research and development direction.

[12] J. S. Katrolia, L. Krämer, J. Rambach, B. Mirbach and D. Stricker, “An Adversarial Training based Framework for Depth Domain Adaptation”, *16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP) 2021*, DOI: 10.5220/0010252403530361.
[13] J. S. Katrolia, L. Krämer, J. Rambach, B. Mirbach and D. Stricker, "Semantic Segmentation in Depth Data: A Comparative Evaluation of Image and Point Cloud Based Methods," *2021 IEEE International Conference on Image Processing (ICIP)*, 2021, pp. 649-653, doi: 10.1109/ICIP42928.2021.9506334.
[14] https://vizta-tof.kl.dfki.de/
[15] https://www.nvidia.com/en-us/self-driving-cars/

5 Related IPR

Not applicable