Deliverable

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

D3.18 Embedded demonstrator AS3 for smart building management

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_D3.18_Public_VF
Responsible beneficiary	DFKI
Contributing beneficiaries	DFKI and IEE S.A.
Contractual delivery date	M42 – 30/10/2022
Actual delivery date	M42 – 25/10/2022
Dissemination level	Public

Document approval

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

Document history

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

Table of content

Executive summary

1 Description of the deliverable objective and content

This deliverable describes the final smart building demonstrator AS3. This report summarizes the hardware and optimized algorithm already described in D3.16 and D3.17, respectively, and describes the work accomplished within the final task 3.D.6 “Algorithm improvement and integration, system testing” to integrate the components and finalize the demonstrator.

[IEE] and **[DFKI]** decided to present the demonstrator in two folds. The first version (see Figure 2) showing the deep learning person detection and segmentation algorithm developed by **[DFKI]** running in real time on a cost-effective embedded platform (see Figure 1), which was one of the main objectives of **[DFKI]** and **[IEE]** within the VIZTA project. The second version of the demonstrator, developed by **[IEE]**, shows effective smart building functionalities: people counting and tailgating/piggybacking detection, the main use-case of the demonstrator (see Figure 3 and Figure 4).

Figure 1: Left: the time-of-flight sensor used on both version of the AS3 demonstrator, the Microsoft Azure Kinect. Right: the low-power consumption platform used on the embedded version of the demonstrator, the Nvidia Jetson Xavier development kit

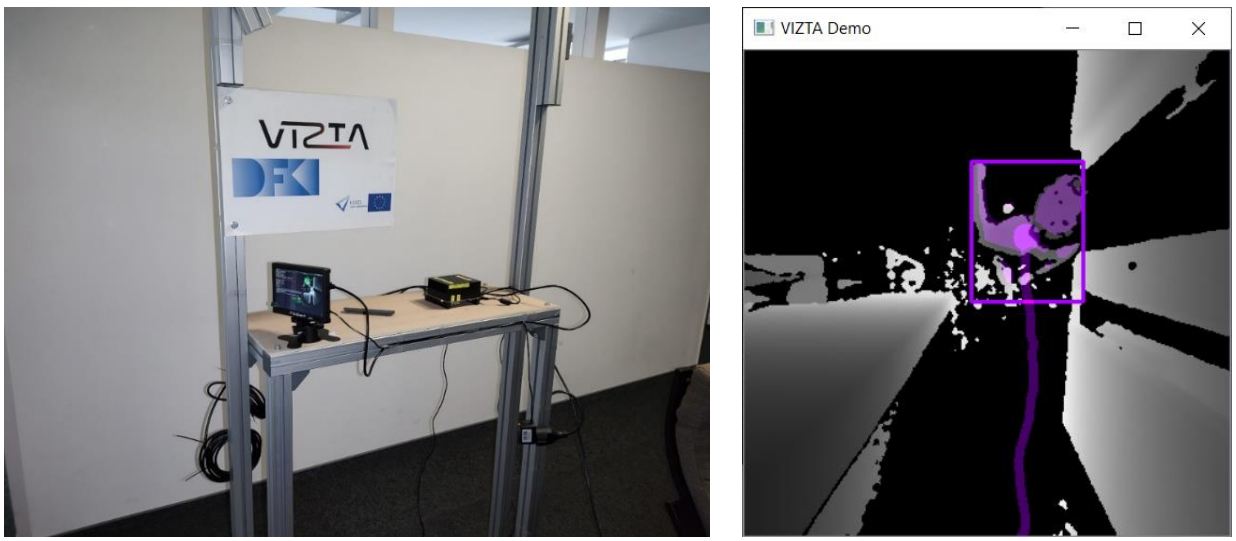

Figure 2: Left: photo of the embedded version of the AS3 demonstrator. Right: demonstration software, running on the Nvidia Jetson Xavier

A major achievement of the research conducted at **[DFKI]** within VIZTA has been the development of an embedded deep learning algorithm for accurate detection and segmentation of persons in high resolution depth images. This optimized algorithm, described in D3.16 has been integrated into the demonstrator and completed during the last period with two additional post-processing functions to increase robustness and enable a tracking. The demonstrator in action will show, as depicted in Figure 2, the detection box and segmentation mask as well as the trajectory of the person in the scene.

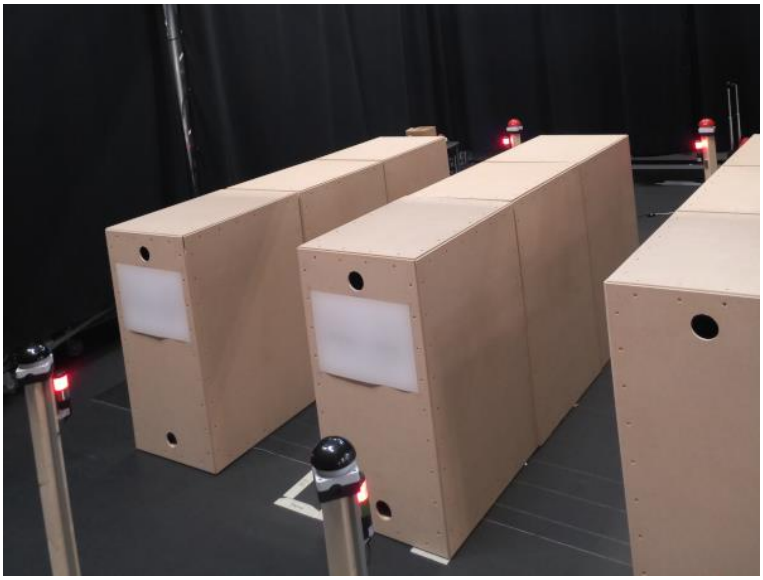

*Figure 3: Picture of the demonstrator setup at **[IEE]** showing the simulated access control gate installed in the field of view of the camera*

In parallel, **[IEE]** has developed a novel high-level building function based on people counting and tailgating/piggybacking detection which is a fast-multi-gate access control. This applicative version of the demonstrator, running on a windows PC, is simulating a two-lane access control, with access requests simulated by push buttons (see Figure 3). A core algorithm performing person detection is encapsulated in an applicative layer performing the people counting functionality as well as the detection of tailgating/piggybacking actions. A dedicated Graphical User Interface allows the demonstrator to display the output of those functionalities: people counts and alarm in case of tailgating/piggybacking detection (see Figure 4).

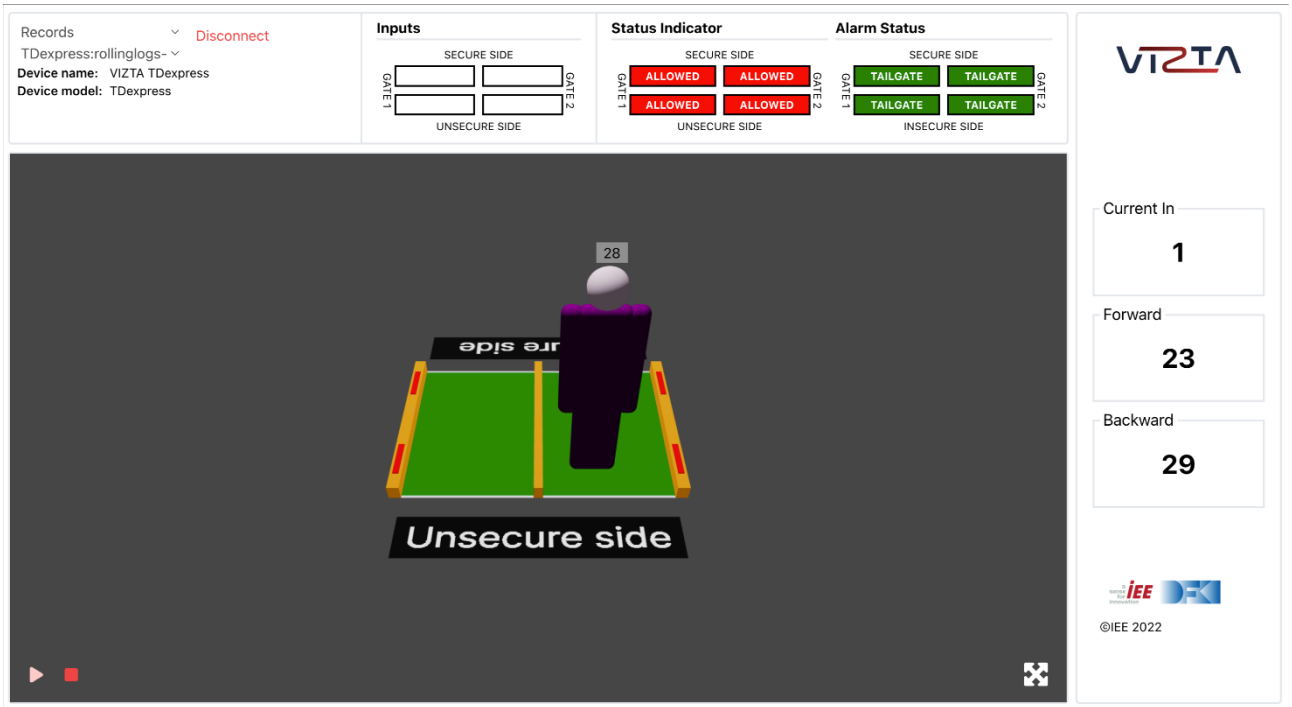

Figure 4: Example of the Graphical User Interface of the applicative final demonstrator showing the different outputs of the applicative software

This report describes the architecture and components of these two variants of the demonstrator and gives an outlook on the exploitation's steps **[IEE]** and **[DFKI]** are intending for next development iterations of the demonstrator towards new products in a research partnership beyond VIZTA.

2 Brief description of the state of the art

Traditionally, surveillance cameras are RGB or IR cameras. For the realization of robust automatic building management functions, time-of-flight depth cameras offer, however, some unique benefits. The depth information they provide allows to detect, classify, and localize persons and objects precisely in 3D space while protecting persons privacy.

For these reasons, the first building management systems based on time-of-flight technology have been developed about 10 years ago[1]. However, as these systems are based on low-resolution time-of-flight sensors, only a few basic building management functions such as people counting, or single access control are realized by these systems.

[1] https://iee-sensing.com/en/building-management-security.html

Future time-of-flight 3D sensors with higher sensitivity and resolution in combination with novel deep-learning algorithms are expected to both enhance the performance of existing building management systems and to realize entirely novel functions as the behaviour analysis of persons and detection of anomalous situations.

In recent years, deep neural network-based approaches have become the state of the art in object detection and are a highly dynamic field of research. Several deep learning methods have been developed specifically for three-dimensional (3D) data for the tasks of point cloud segmentation and shape classification[2]. As a basis for the task of real-time object detection and segmentation in high-resolution depth images, these methods are, however, too memory and computation time demanding.

The preferred approach for a real-time object detection in 3D data is therefore to adapt state-of-the-art 2D convolutional networks to depth data. These detector approaches can be broadly divided into two categories based on their architecture. More accurate object detectors employ a 'propose regions first, then detect' approach, where several regions are first selected from the image based on their likelihood to contain an object. These regions are used to pool features within them and then to compute the location of the object and the probability that they belong to a certain class. Architectures like R-CNN[3], Fast R-CNN[4] and Faster R-CNN[5] fall in this category and are also known as two-stage detectors.

The second class of object detectors are one-stage detectors, which include faster but less accurate architectures. They treat object detection as a problem of regression. The objective is to directly arrive at the bounding box coordinates and the class probabilities of the objects from the feature maps extracted from the image. These single stage detectors like YOLO[6], YOLACT[7] and SSD[8] have real-time detection speed and comparable accuracy to two-stage detectors.

---

[2] Guo, Yulan & Wang, Hanyun, & Hu, Qingyong, & Liu, Hao & Liu, Li, and Bennamoun, M. Deep Learning for 3D Point Clouds: A Survey, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), Early access paper 2020

3 Girshick, Ross & Donahue, Jeff & Darrell, Trevor & Malik, Jitendra. (2015). Region-Based Convolutional Networks for Accurate Object Detection and Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence. 38. 1-1. 10.1109/TPAMI.2015.2437384

4 Girshick, Ross. (2015). Fast r-cnn. 10.1109/ICCV.2015.169.

5 Ren, Shaoqing & He, Kaiming & Girshick, Ross & Sun, Jian. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence. 39. 10.1109/TPAMI.2016.2577031

[6] Redmon, Joseph & Divvala, Santosh & Girshick, Ross & Farhadi, Ali. (2016). *You Only Look Once: Unified, Real-Time Object Detection*. 779-788. 10.1109/CVPR.2016.91

[7] Bolya, Daniel & Zhou, Chong & Xiao, Fanyi & Lee, Yong. (2019). YOLACT: Real-time Instance Segmentation

[8] Liu, Wei & Anguelov, Dragomir & Erhan, Dumitru & Szegedy, Christian & Reed, Scott & Fu, Cheng-Yang & Berg, Alexander. (2016). *SSD: Single Shot MultiBox Detector*. 9905. 21-37. 10.1007/978-3-319-46448-0_2

For the task of 2D object segmentation Deep Convolutional Neural Networks initially intended for image classification like Resnet[9] have been modified into a fully convolutional form to adapt them for the task of semantic segmentation[6][10,11]. A state-of-the art method is Mask R-CNN[11], an extension of the two-stage detector Faster R-CNN which incorporates a parallel mask prediction branch. In the same vein, the single stage detector YOLACT[6] also includes an additional branch to predict masks along with boxes. This aspect together with the run-time benefit of a single stage detector makes YOLACT particularly interesting for a real-time application. Nevertheless, to deploy a neural network like YOLACT on embedded platform dedicated for AI application, the algorithm had to be adapted and optimized, for the dedicated embedded platform.

3 Deviation from objectives and corrective actions

The demonstrator AS3 is equipped with an Azure Kinect 2D/3D camera which had been used as development platform for this demonstrator throughout the whole project. Originally, it was planned to integrate the S2 evaluation kit delivered by **[ST GNB2 SAS]** into the final demonstrator. However, as already reported in D3.17, the assessment of this sensor showed that it does not fulfil the requirements in terms of range and field-of-view for the demonstrator AS3 without major changes which were not feasible within VIZTA project. Therefore, the contingency plan (see WP3, risk IDs #13 and #14) had been activated per common agreement between **[ST GNB2 SAS]**, **[DFKI]** and **[IEE]** to use the preliminary camera Azure Kinect for the final demonstrator AS3.

4 Impact of the results

One version of the demonstrator combines a high-resolution time-of-flight sensor with an energy and cost-efficient embedded processing platform dedicated for AI applications. A novel deep-learning algorithm person detection and segmentation algorithm developed at **[DFKI]** within VIZTA is deployed on this processing platform for online demonstrations. The impact of this embedded demonstrator for the development of novel building management functions is manifold.

The novel deep learning algorithm and the embedded processing platform of the demonstrator can be integrated in existing building management systems for people counting or tailgate detection to enhance their performance or extent the functionality. Within VIZTA **[IEE]** has already

[9] He, Kaiming & Zhang, Xiangyu & Ren, Shaoqing & Sun, Jian. (2016). *Deep Residual Learning for Image Recognition*. 770-778. 10.1109/CVPR.2016.90.
[10] Chen, Liang-Chieh & Papandreou, George & Kokkinos, Iasonas & Murphy, Kevin & Yuille, Alan. (2016). DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence. PP. 10.1109/TPAMI.2017.2699184.
[11] He, Kaiming & Gkioxari, Georgia & Dollár, Piotr & Girshick, Ross. (2017). Mask R-CNN

developed a novel higher-level functionality of a fast-multi-gate access control system which is shown in the second version of the AS3 demonstrator.
This is one functionality which has the potential to profit from the availability of high resolution ToF sensors and corresponding detection algorithms in terms of scalability and speed. Partners **[IEE]** and **[DFKI]** are already evaluating the integration of the **[DFKI]** embedded algorithm in the **[IEE]** application demonstrator and intend to cooperate after VIZTA in feasibility studies concerning future products of **[IEE]** smart buildings portfolio.

In addition, the embedded demonstrator can serve as a basic to realize entirely novel functions as the behaviour analysis of persons or detection of anomalous situations. First research results in this direction have already been published[12] as a result of VIZTA by **[DFKI]** and the publicly available benchmark dataset *VIZTA-Timo*[13] created within VIZTA is fostering research in this direction. This dataset hosted on **[DFKI]**s webpage as well as the scientific publications related to it and the algorithm research done for VIZTA will increase the attention on **[DFKI]**'s expertise and know-how in the development of leading-edge embedded deep learning algorithms. Therefore, a positive impact on the acquisition of new research projects and partnerships with industry is expected, especially as the develop embedded detection and segmentation algorithm is applicable in many industry files, as industry automation, automotive, robotics or construction industry.

5 Related IPR

Not applicable

[12] Pascal Schneider; Jason Raphael Rambach; Bruno Mirbach; Didier Stricker, *Unsupervised Anomaly Detection from Time-of-Flight Depth Images*, CVPR Workshop on Perception Beyond the Visible Spectrum, 2022
[13] https://vizta-tof.kl.dfki.de/timo-dataset-overview/