JDST, vol. 5, no. 3, p44-p57, 2022
https://doi.org/10.46713/jdst.05.003

www.jdst.eu, ISSN 2534-9805, e-ISSN 2534-9813

Research Article

Novel Smart Sensor Technology Platform for Border Crossing Surveillance within FOLDOUT

Cristina Picus, Daniel Bauer, Michael Hubner, Andreas Kriechbaum-Zabini, Andreas Opitz, Julian Pegoraro, David Schreiber, Stephan Veigl

AIT Austrian Institute of Technology GmbH, Giefinggasse 4, A1210 Vienna, Austria, www.ait.ac.at

ABSTRACT

In this paper we introduce AIT's Smart Sensor Technology platform, developed within the FOLDOUT project (Through-foliage detection, including in the outermost regions of the EU). The novel platform is part of FOLDOUT's solution for ground-based border surveillance, and in particular detection of foliage penetration. The platform combines a high resolution (4K) RGB camera, a thermal camera, as well as a 30x zoom (2K) NIR low-light camera. It is designed to work day and night under strongly varying weather conditions, it is transportable and can be deployed in remote areas, as well as being self-sufficient if operated by battery. The typical sensor range is about 100-200m which can be adapted per specific usage. The platform might be used for surveillance of dedicated areas or by providing an open interface to be integrated into whole surveillance systems in which ground-based sensors are combined with high altitude sensors to provide both close and wide area coverages. The platform provides state-of-the-art AI-based detection algorithms for object detection and classification, for both RGB and thermal images. The individual detections are further fused by establishing coincidence in time and space between the individual detectors, where sensor is geo-localized for accurate target localisation and visualisation on a map of the area to be monitored.

✉ Corresponding Author: Cristina Picus; Tel.: ++43 664 8251099;

E-mail: cristina.picus@ait.ac.at

ARTICLE INFO

RECEIVED: 09 Oct 2021
REVISED: 10 Nov 2021
ACCEPTED: 30 Nov 2021
ONLINE: 12 DEC 2021

KEYWORDS

AI, border surveillance, machine learning, occlusions, sensor fusion, visual sensors

I. INTRODUCTION

Surveillance of remote areas between regular border crossing points is one of the main tasks of border guards operating at the EU/Schengen's external borders. Most of the land borders are equipped with standard surveillance technologies, such as cameras, IR sensors, seismic sensors, etc., to monitor activities in areas where border crossings are not normally foreseen. However, coverage is not always possible for areas that can include diversified landscape: from soil fields to forests, various types of vegetation, altitude differences and difficult to access terrain. In addition, the available technologies have clear limitations when it comes to surveying areas such as forests or areas with dense vegetation, which provide easy hiding places for potential underlying irregular activities.

Moreover, facilitators of illegal activities are often aware of the stationary installed technologies and quickly learn how to avoid such areas. Furthermore, routes and activities are often contingent upon the circumstances and new routes may temporarily become relevant, e.g., in case of crisis, emergencies etc.

In particular, foliage penetration is an unsolved important part of border surveillance. By solving the problem of unreliable detections in such harsh environments, border guards' workloads are reduced. Detecting penetrating objects (e.g., persons and cars) through dense foliage in various climate conditions using visual sensors is prone to high fault rates. The project FOLDOUT builds a system that combines various sensors and technologies and fuses these into a robust detection platform. In this paper, we focus on the work done within FOLDOUT concerning visual sensors. We have identified the need for novel visual sensors to improve the situational awareness at the border, sensors which are: 1) deployable in remote forested areas with minimal infrastructure, 2) able to operate 24/7 under harsh weather and environmental conditions, and 3) coping with limited visibility due to dense foliage coverage. Furthermore, there is an additional need for ad-hoc flexible sensors that can be easily integrated in existing surveillance infrastructures to improve coverage or to react to unforeseen events.

This paper presents a novel sensing platform that addresses all these aspects, namely the AIT's Smart Sensor Technology platform. The solution is a standalone platform integrating different complementary visual sensors and targeting ground-based surveillance. The sensor combines a high resolution (4K) RGB camera, a thermal camera, as well as a zoom (2K) NIR low-light camera. It is designed to work day and night under strongly varying weather conditions. It is transportable and can be deployed in remote areas, as well as potentially self-sufficient if operated by battery. The typical sensor range is about 100-200m which can be adapted per specific usage. It can be used for surveillance of dedicated areas or providing an open interface to be integrated into surveillance systems in which potentially ground-based sensors are combined with high altitude sensors to provide both close and wide area coverages.

This is the targeted usage within the H2020 project FOLDOUT (Through-foliage detection, including in the outermost regions of the EU) for which the above sensor (AIT'S Smart Sensor Technology platform) has been originally developed [1]. The sensor provides advanced detection functionalities by novel AI-based algorithms for object detection and classification under partial occlusion such those originating by objects partially hidden under vegetation. The sensor is automatically geo-localized for accurate target localisation and visualisation on a map of the area, useful to provide the operator with visual feedback in a GUI and in potential combination with further sensors deployed in the area to be monitored.

The rest of this paper is organized as follow. In section II. STATE-OF-THE-ART our novel Smart Sensor Technology platform is compared with existing state-of-the-art sensor solutions. Section III. AIT's Smart sensor Technology platform solution describes the details of our platform solution, including its technical specifications (A. Technical specifications), architecture (. Software and communication), AI-based detection algorithms (C. AI algorithms), and the geo-registration (D. Geo-Registration and Fusion). Section IV. CONCLUSIONS contains our conclusions.

II. STATE-OF-THE-ART

COMPARISON WITH AVAILABLE PRODUCTS

Parameter/Product	AIT (Smart Sensor)	Exensor (Scout Mk3)	Dahua (ACP)	Flir (Triton)	Axis (Q87)	Hikvis. (DS-2TD6266)
Platform setup	Lense options: Visual/Thermal, D-N, LTE, IMU/GPS, IR	Lense Options	2 Lense Options			Lense Options
Edge Device/ Analytics	AI Object & motion detection (Visual & Thermal)	Minimal (motion detection)		minimal	minimal	Minimal
Resolution (Visual/ Thermal), Zoom	12MP, VGA, D-N zoom	752BW, VGA	2MP, VGA, Vis/Th zoom	1MP, VGA, Vis/Th zoom	1MP, VGA, Vis/Th zoom	1MP, VGA,
Transport. Design	Yes	Yes				
Open Interface	Yes		Yes	Yes	Yes	Yes
Other	Sensor Fusion, calibration, detection range 200m, customized solution	Only with PIR system, detection range 50m	Detection range 3km, IVS, heavy	P/T, VMS	P/T, detection range 3km	P/T, IR 500m
Costs – L (<25k€), H (>25k€)	L	L	HH	H	L-H	L-H

Figure 1: Main features comparison between AIT's Smart Sensor Techonolgy platform and available state-of-the-art solutions on the market

Although border control units are typically well equipped with state-of-the-art surveillance technology, current equipment has various drawbacks, which can be summarized as follows:

- Low performance in difficult outdoor environments, e.g., foliated areas;
- Low performance of event detection (in particular, high false alarms rate);
- Difficult deployment in remote areas;
- Different solutions required for different environment scenarios;
- High operational costs – scalability and interoperability are limited;
- Difficult to use and deploy by the operators;

The platform presented in this paper offers various advantages over existing solutions, hence overcoming current limitations. It is designed to be deployed in remote areas, with high mobility even on challenging terrain and forested areas. It provides an improved automatic event detection targeting also through-foliage detection of partially occluded objects. Furthermore, it has an open interface to be integrated into existing surveillance applications, with consequent cost reduction for operation and improved usability by the operator. There are alternatives on the market, that target similar applications, most of them are however meant for stationary use (see comparison with existing solutions in Figure 1).

Dahua [2] builds some combined bi-spectral cameras with integrated video analytics, but these are very heavy – up to 100Kg – and require heavy structures for mounting and are not meant to be moved and deployed in different terrains. FLIR has some few products with integrated analytics, which are also not meant to be moveable [3] and the HIKVISION 2TD6266 series [4] is produced for fixed installations. The above-mentioned cameras are meant targeted for perimeter surveillance in fixed installments and belong to the high-price sector. An additional platform which provides similar functionalities is the Exensor Scout MK3 [5], however only very basic analytics (motion detection) is provided and with low resolution RGB sensor. The Axis Q87 series [6] combines as well bi-spectral cameras, but without fully customizable integrated analytics.

III. AIT's Smart sensor Technology platform solution

The Smart Sensor Technology platform (see Figure 2, left and Figure 3) targets day and night vision analysis by combining a high resolution RGB camera, a thermal camera, and a zoom NIR low-light camera. The combination of different lenses provides an all-in-one solution for day/night operation, which can capture high quality images even in challenging conditions and dark environment. Currently, the platform targets distances up to 200m for visual surveillance of large outdoor areas. It has integrated GPS and IMU sensors that enable self-localization and recognition of camera orientation. The output consists of 3 video streams, each including camera position and orientation. The platform has built-in analytics functionalities targeting detection and classification of different type of targets such as persons, vehicle etc. Independent detectors based on deep neural networks perform video analytics for object and event detection on each video stream. The platform integrates an NVIDIA Jetson AGX Xavier board [7] including a CPU as well as integrated GPU units, for onboard data processing. This enables sending over the wireless network only the detection results rather than the raw data, both to avoid overloading of the network as well as to fulfill basic privacy by design requirements. Streaming is performed only upon explicit request by the operator. The platform

Figure 2: Left: The Smart Sensor Technology platform; Right: Back panel components and cable connections

can be optionally operated by battery, which enables autonomy for deployment in remote environments.

The main features of the platform can be summarized as follows:

- Multi sensor detection system for usage in harsh environments;
- Complementary sensors for high performance under specific settings: 4K RGB, thermal, 2K low-light NIR zoom camera;
- Zoom option as an additional feature for the operator;
- Designed for day/night, harsh weather conditions;
- Configurable for different distances;
- Transportable and robust;
- Simple deployment - mountable on trees or masts;
- GPS and IMU for self-localization and estimation of camera orientation;
- 4G LTE, WLAN/Wi-Fi, Bluetooth communication on board;
- Self-sustaining with battery option;
- Optionally, a visual or infra-red flood light can be installed on the pan-tilt mount;

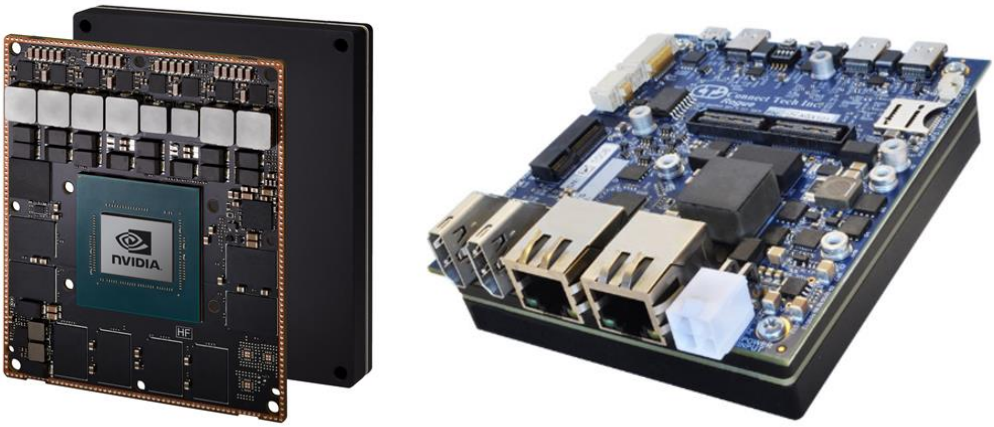

Figure 3: Embedded system with NVIDIA Jetson AGX Xavier board

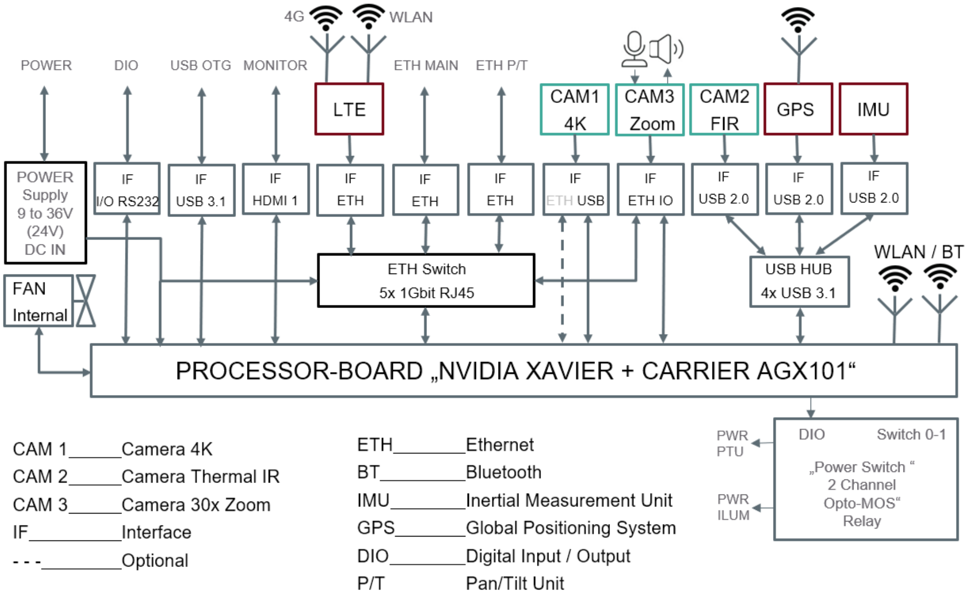

Figure 4: Block diagram electrical architecture

A. Technical specifications

The main features of the Smart Sensor Technology platform are (see Figure 2, right and Figure 4):

- High quality 4K RGB imaging, 4000x3000p resolution (FLIR);
- Thermal camera, 640x480p resolution (FLIR);
- 30 x optical zoom 2K NIR camera, 1920x1080p (SONY);
- Pan-tilt unit (FLIR), with panning range of [-168°, +168°] and tilting range of [-90°, +30°].
- Video streaming RTSP;
- Dimensions: 35cm x 22cm x 9cm (without the pan-tilt unit);
- Weight: 6kg for the whole platform; pan-tilt unit adds another 6.5kg;
- Ruggedized, passive cooled, water resistant housing for multiple sensors & optional IR light support;
- Onboard processing with Nvidia Jetson Xavier;
- 2TB SSD fast NVME storage (expandable via 2nd M.2 connector on Jetson Xavier);
- Connectivity 1Gbit Ethernet, Bluetooth, WLAN/WI-FI, 4G LTE;
- I/O support on Jetson Xavier USB 3.0 (OTG), HDMI, 2x RS232, CAN-Bus, DIO (Optional);
- GPS, 9-axis IMU;

- Power switches for pan-tilt unit and IR light;
- Mountable on pan-tilt unit or stand alone;
- Power: 12-30VDC cable or battery;
- Maximum power dissipation 75Watt (without pan-tilt unit and IR light), depending on the algorithm processed and the Jetson Xavier power modes;
- Optional pan-tilt unit support via zoom camera module (RS485, Pelco D/P);
- Optional Analog Audio In/Out and Video Out via zoom camera module;

B. Software and communication

The software of the Smart Sensor Technology platform is based on a modern, open and modular architecture. It is comprised of two tiers of communication/interfaces layers. The internal camera implementation utilizes the Robot Operating System (ROS) [8] to provide a service and communication framework. Due to its flexible and modular architecture, ROS is well-known and widely used by the robotics community but also many other industries. The communication protocol is based on ROS messages or Common Alerting Protocol (CAP). CAP is a standardized format for alerting, which has been further specified within the FOLDOUT project to cover dedicated layers of the alerting system, which include: raw detection by the different sensors, registration of detections on a global map, fusion of the detections, tracking the fused detections, analysis of the tracks, and raising an alarm when needed.

The external interface utilizes the Foldout CAP/CAL communication, where CAL stands for Communication Abstraction Layer. From the Foldout point of view, the Smart Camera is treated as a black-box sensor. It registers itself as camera sensor at the Foldout CAL server. Besides the sensor type, additional meta data of the platform, a URL to the video stream, the current position and orientation, as well as a periodic heartbeat are sent to the server. Detections in the sensors are converted to Foldout events and sent in extended CAP packages. All external communication with the Foldout system is handled by one interface node implementing the Facade software design pattern [9] shielding all internal details from the outside. Due to the utilization of a single node for all external communication only one node has to be altered to adapt the camera to a new environment.

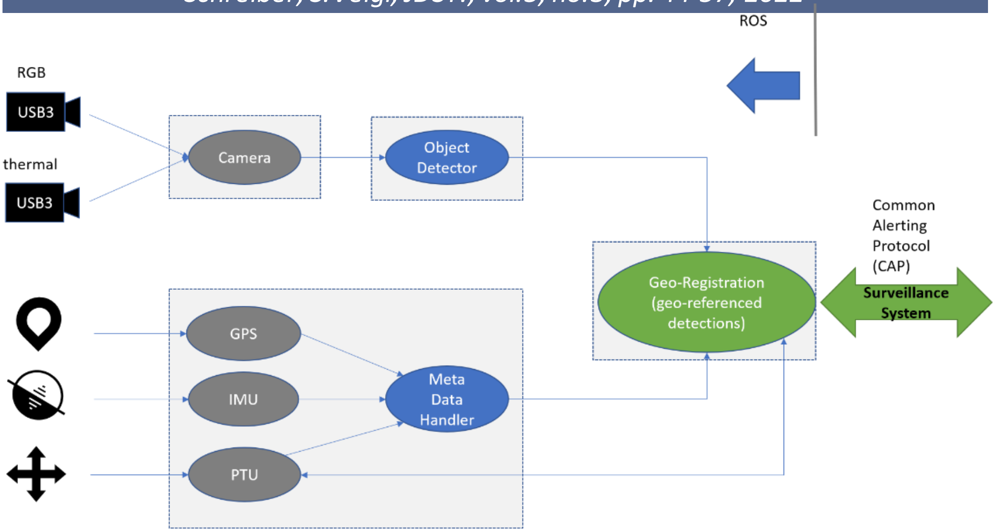

Figure 5: Modules of the Smart Sensor Technology platform

Each of the hardware components and functions are wrapped by individual software modules [9] providing an abstraction layer and low-level interface. The low-level modules are designated as gray modules in Figure 5. The next layer is the computation layer (designated in blue in Figure 5), also often called as business logic layer. This layer contains main computer vision algorithm (object detection) as well as the Metadata Handler. The Metadata Handler module collects the information of the low-level sensor modules and generates a single metadata package for the smart platform. The final layer, the interface layer, consists only of one Facade module that can be customized to the specific interface of the project and different environments. The detection results are enriched at this level with the geo-position of the camera. With this information the geo-localization of a detection can be computed by the external consumer system of these data. Further, at this level, a fusion of the detections from different sensors can be done. Depending on customization, either ROS, CAP any proprietary message (with an according interface node) can be sent. In the context of the project FOLDOUT, the ROS messages are converted to the Foldouts CAP flavor format.

For steering the Smart Sensor Technology platform, a connection to the pan-tilt-unit is established via an ethernet ASCII interface. The required commands such as getting and setting pan and tilt position are wrapped in a ROS node as custom services with a corresponding publisher. Additionally, the possibility to power cycle the device via the Jetson's GPIO (General Purpose Input and Output) interface was included in the ROS interface. In case that the Smart Sensor technology platform is installed without a pan-tilt-unit, the readings from the inertial measurement unit (IMU) can be used for obtaining pitch, roll and yaw. This is done by querying the corresponding ROS interface wrapped around the device's virtual COM port interface. The published ROS messages are then processed by the Metadata Handler ROS node where small measurement errors are filtered out.

The GNSS position information is obtained using a GNSS receiver module and an external antenna. The position is published as a ROS message and, similarly to the IMU, processed by the Metadata Handler which filters out fluctuating measurement values. All sensor interfaces described above are wrapped, deployed, and executed as Docker containers [10]. This ensures a reproducible execution environment and enables easier handling of the individual containers via the docker-compose management interface.

C. AI algorithms

Foliage penetration is an unsolved important part of border surveillance. By solving the problem of unreliable detections in such harsh environments, border guards' workloads are reduced. Detecting objects (e.g., persons and cars) through

Figure 6: Typical detection results displayed as bounding boxes for RGB (left) and thermal images (right)

dense foliage, in various climate conditions, using visual sensors is prone to high fault rates. Due to the dense foliage, FOLDOUT scenarios contain an unprecedent amount of occlusion - in fact fragmented occlusion (for example, looking through the branches of a tree). Fragmented occlusions are much more challenging than ordinary partial occlusions as typically only small fragments of the persons are seen through the foliage. Currently, state-of-the-art person detectors are trained with datasets containing only partial occlusions.

By using the integrated GPU cores of the Smart Sensor Technology platform, Deep Learning-based state-of-the-art object detectors can run directly on the platform. The range of detection depends entirely on the camera and optics used. For FOLDOUT needs, detection range is up to 200 meters. The Smart Sensor Technology platform uses both the RGB and the thermal cameras for detection. Furthermore, in previous work [11] we have shown that in a controlled environment, classification with hyperspectral imaging is also possible. A state-of-the-art detector (YOLOv5) [12] is used for FOLDOUT current requirements, but other models can be executed as well for specific purposes. The YOLOv5 model used has been trained on the MS COCO dataset [13] and is therefore trained to detect 80 classes. Figure 6 shows typical detection results (displayed by bounding boxes) for RGB (left) and thermal (right) images.

In a previous FOLDOUT publication [14], an attempt was made to train a state-of-the-art Deep-Learning-based detector (specifically, Mask R-CNN detector) on new training data which captures explicitly the problem of fragmented occlusions. The results show some improvements of the Mask R-CNN detector with this new training strategy (also against other detectors) for data showing slight fragmented occlusion, however less improvement was gained for the case of heavy fragmented occlusions. Consequently, an additional detection algorithm has been tested, which is based on a background learning algorithm (an approach for detecting moving objects in videos based on changes in the learned background image). We have developed a robust real-time background modeling algorithm which is able, on one hand, to detect moving persons even when they are heavily occluded by foliage, while, on the other hand, being resilient against dynamic background such as moving foliage and consequent lighting changes.

The background modeling approach is pixel-based and comprises three modules. The core module is similar to a previous work of ours, published in [15], namely using an efficient non-parametric representation of the history of each pixel. The second, post-processing module, refines the foreground detection of the first module by carefully removing foreground which is due to dynamical background, e.g., moving vegetation and corresponding lighting changes. Finally, the third module is comprised of an event detector, which makes use of smoothed local spatial-temporal regions of the foreground mask to predict the presence of moving persons (an event detector). The background-modeling-based algorithm works with both the 4K-RGB and the thermal cameras.

shows an example of background-based event detection on an RGB frame, with several persons partially and heavily occluded behind the trees.

Figure 7: RGB frame (left) and the corresponding detections of occluded persons by the background-based detector (Right)

D. Geo-Registration and Fusion

The Smart Sensor Technology platform comprises two independent detectors, namely a dedicated detector per camera (RGB and thermal). To enhance accuracy and minimize false positives of the individual detectors, a feature-based fusion approach was developed which can be integrated into the platform. The fusion approach consists of the following two main steps: ***Geo-Registration****:* Transforming the detections (Bounding Boxes in the video frames) into a global coordinates frame, and ***Fusion:*** Establishing coincidence of multi sensor detections in time and space using probabilistic occupancy grids [16] and a Linear Opinion Pool [17] method to combine the maps of the individual sensors.

To perform the geo-registration, it is essential, that the Smart Sensor Technology platform is calibrated intrinsically (cameras parameters) as well as extrinsically (orientation and position of the platform). For this purpose, the inertial measurement unit (IMU) and a GPS module were integrated into the platform.

The fusion approach utilizes weighted maps (inspired by the probabilistic occupancy maps concept), where for each individual camera, a dedicated weighted map is generated, representing the likelihood for a true detection within each grid cell of the map. The stream of events from both cameras are used to update the corresponding maps over time. A weighted map yields high weight values if the corresponding camera provides many events at the same location in a short time span. The fusion map is computed as a linear combination of the 2 individual maps, next a set of areas with potential alarm are generated, via segmentation of the fusion map. The polygons enclosing these areas can be sent and displayed to an operator on a map of the surveillance area.

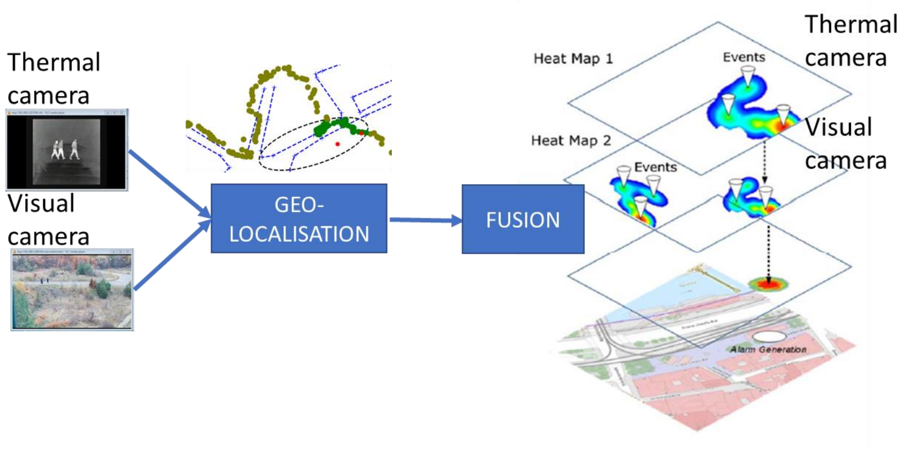

Figure 8: Geo-localisation and fusion. The information from thermal and visual cameras is geo-localized and aggregated into multiple Heat Maps to achieve coincidence in time and space.

IV. CONCLUSIONS

The article describes the novel Smart Sensor Technology platform developed by AIT within the H2020 funded project FOLDOUT. The camera offers clear advantages with respect to commonly available surveillance cameras, especially for what concerns portability and easy integration to an existing surveillance infrastructure. The platform combines high resolution and thermal cameras, as well as low-light camera with zoom functionality. The software is based on an open and modular architecture enabling easy interfacing to external systems. Further, it provides advanced AI-based algorithms targeting detection of different type of objects, also under partial occlusion, such that occurring in areas with vegetation. Finally, an automatic geo-localization of the sensor enables an accurate localization of the targets and corresponding visualization in a global map for the operator. It can also be additionally used to fuse detections from the RGB and thermal cameras.

ACKNOWLEDGEMENTS

The work described in this article was accomplished within the FOLDOUT project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 787021.

References

[1] "The Horizon 2020 Programme FOLDOUT project," [Online]. Available: https://foldout.eu/.

[2] "Dahua Security," [Online]. Available: https://www.dahuasecurity.com/products.

[3] "FLIR," [Online]. Available: https://www.flir.com/.

[4] "HIKVISION," [Online]. Available: https://www.hikvision.com/de/.

[5] "Exensor," [Online]. Available: https://www.exensor.com/products/flexnet-flexible-network-of-sensors/electro-optics/scout-mk3/.

[6] "Axis," [Online]. Available: https://www.axis.com/products/axis-q87-series.

[7] "NVIDIA Developer Blog," [Online]. Available: https://developer.nvidia.com/blog/nvidia-jetson-agx-xavier-32-teraops-ai-robotics/.

[8] "The Robot Operating System (ROS)," [Online]. Available: https://www.ros.org/.

[9] E. Gamma, R. Helm, R. E. Johnson and J. Vlissides, in *Design patterns: Elements of reusable object-oriented software*, Addison-Wesley Longman Publishing Co., Inc., 1995.

[10] "docker," [Online]. Available: https://www.docker.com.

[11] A. Papp, J. Pegoraro, D. Bauer, P. Taupe, C. Wiesmeyr и A. Kriechbaum-Zabini, „Automatic Annotation of Hyperspectral Images and Spectral Signal Classification of People and Vehicles in Areas of Dense Vegetation with Deep Learning,“ *Remote Sensing,* том 12, № Remote Sensing 12.13 (2020): 2111., 2020.

[12] G. Jocher, K. Nishimura, T. Mineeva and R. Vilariño, "YOLOv5 (2020)," [Online]. Available: https://github.com/ultralytics/yolov5.

[13] T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*, 2014.

[14] J. Pegoraro and P. R., "The Problem of Fragmented Occlusion in Object Detection," in *Joint Austrian Computer Vision and Robotics Workshop*, 2020.

[15] D. Schreiber and M. Rauter, "GPU-based non-parametric background subtraction for a practical surveillance system," in *12th IEEE International Conference on Computer Vision (ICCV) Workshop on Embedded Computer Vision*, Kyoto, Japan, 2009.

[16] J. Tong, L. Chen and Y. Cao, "Human positioning based on probabilistic occupancy map," in *14th IEEE International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery*, Huangshan, China, 2018.

[17] J. D. Adarve, M. Perrollaz, A. Makris and C. Laugier, "Computing occupancy grids from multiple sensors using linear opinion pools," in *IEEE International Conference on Robotics and Automation*, St Paul, MN, USA, 2012.