



A Cyber-Physical System for Integrated Remote Control and Protection of Smart grid Critical Infrastructures

Juan Isern¹ · Gabriel Jimenez-Perera¹ · Luis Medina-Valdes² · Pablo Chaves³ · David Pampliega³ · Francisco Ramos³ · Francisco Barranco¹

Received: 1 April 2021 / Revised: 31 March 2022 / Accepted: 16 January 2023 / Published online: 22 February 2023
© The Author(s) 2023

Abstract

This work proposes a Cyber-Physical System (CPS) for protecting smart electric grid critical infrastructures using video surveillance while remotely monitoring them. Due to the critical nature of the smart grid, it is necessary to guarantee an adequate level of safety, security and reliability. Thus, this CPS is back-boned by a Time-Sensitive Network solution providing concurrent support for smart video surveillance and smart grid control over a single communication infrastructure. To this end, TSN delivers high-bandwidth communication for video surveillance and deterministic quality of service, latency and bandwidth guarantees, required by the time-critical revision smart grid control. On the one hand, the CPS utilizes High-availability Seamless Redundancy in the control subsystem via Remote Terminal Units (RTUs) guaranteeing seamless failover against failures in smart grid. On the other hand, the smart video surveillance subsystem applies machine learning to monitor secured perimeters and detect people around the smart grid critical infrastructure. Moreover, it is also able to directly interoperate with RTUs to send alarms in case of for example, an intrusion. The work evaluates the accuracy and performance of the detection using common metrics in surveillance field. An integrated monitoring dashboard has also been developed in which all CPS information is available in real time.

Keywords Smart grid · Time sensitive network · Smart video surveillance · Redundancy

1 Introduction

The far-reaching EU Security Union Strategy includes energy infrastructures: individuals rely on key infrastructures in their daily lives, to travel, to work, to benefit from essential public services such as hospitals, transport, energy supplies, or to exercise their democratic rights. If these infrastructures are not sufficiently protected and resilient, attacks can cause huge disruptions (physical or digital) [1].

Over the last decades, the requirements for the stability and reliability of the power supply have increased with the continuous development of our society [2]. Control and monitoring of the electric network is of great importance to secure the stability of the electricity supply and to guarantee the needs of citizens and industry. Such control of critical infrastructures is a task of great complexity due to the risks that the security personnel need to supervise. Investing only on preventive equipment and security personnel resources is insufficient; autonomous surveillance systems that improve the protection and reduce cost are needed [3]. Preferably, integrated systems that enable seamless interoperation across the subsystems in electrical substations and allow for joint strategies are the aim of future solutions [4].

✉ Juan Isern
jisern@ugr.es

Gabriel Jimenez-Perera
gabrieljimenez@ugr.es

Luis Medina-Valdes
luis.medina@sevensols.com

Pablo Chaves
pablo.chaves@se.com

David Pampliega
david.pampliega@se.com

Francisco Ramos
francisco.ramos@se.com

Francisco Barranco
fbarranco@ugr.es

¹ Computer Architecture and Technology CITIC, University of Granada, Granada 18014, Spain

² Seven Solutions, Granada 18014, Spain

³ Schneider Electric España S.A., Sevilla 41092, Spain

The main elements of the electrical distribution network are the Substation Automation Systems (SAS) [5] which monitor and control the electrical infrastructure. Distribution networks comprise from distribution substations to the service entrance of the electricity consumers, including distribution substations, primary feeders, distribution transformers, and secondary systems [6]. A substation is a high-voltage electric system facility used to switch generators, equipment, and circuits or lines in and out of the system. It is also used to change voltage levels or to switch between alternating and direct current. A key element of the SAS are the Remote Terminal Units (RTUs) which play an important role in the monitoring and control of the infrastructure deployed at those substations. Communications are also essential for a proper operation, and redundant technologies such as High-availability Seamless Redundancy (HSR) contribute to the reliability by decreasing probability of communication failures [7].

While intelligent video analytics is the most widely used technology in safety and security [8], the use of this technology is not widely deployed in electric substations in general, or in final distribution substations in particular [9]. Non-scheduled service interruptions come at a significant cost, both economic and of reputation, positioning supply, quality, reliability and cost penalties at the forefront of interests for utilities [10]. Therefore, the electric companies benefit from designing and building their substations with built-in video surveillance systems [2, 11]. However, this continuous online monitoring produces vast volumes of data that need to be timely analyzed in an efficient manner, while avoiding the interference with control and monitoring [12].

The first part is achieved thanks to powerful and efficient embedded devices, part of the data is locally processed, reducing bandwidth usage and latency specially in isolated or limited connection scenarios. However, the rest of data analysis that gathers local data is usually offloaded to high-performance machines [13].

The second part is achieved by establishing data traffic priority and guaranteeing specific bandwidth usage to the critical data. Having an integrated and stable communication flow between the video surveillance subsystem and the electrical substation control subsystem reinforces and increases the safety and security level of smart grid critical infrastructures. Time Sensitive Networking (TSN) is a set of extensions of the IEEE 802.1 and 802.3 standards adding deterministic Quality of Service (QoS) to bridged Ethernet networks, such as bounded latency and guaranteed bandwidth [14]. The well-known interoperability between different applications and equipment provided by Ethernet is extended to integrate conventional best-effort data flows with hard real-time communication on one network infrastructure. This deterministic QoS is ultimately supported by sub-millisecond time synchronization.

The aim of this work is to propose an active video-surveillance CPS for the prevention of potential harms to critical infrastructures such as the smart electric grid. Therefore, the main contributions are: 1) the integration of a bandwidth-intensive subsystem such as video surveillance sharing the same TSN with smart grid traffic, without compromising the operation of both subsystems; 2) real-time video processing and automation of surveillance of smart grid critical infrastructures, using distributed computing between local edge nodes and a central cloud server; 3) the development of a video surveillance pipeline tested with a state-of-the-art benchmark dataset and whose GPU-intensive tasks embedded for local processing at the edge with similar results; 4) the integration of alarms from video surveillance and the SAS, making it possible to automatically perform actions to protect people in the substation and equipment accordingly; 5) the implementation of a dashboard displaying real-time heterogeneous information from the three different subsystems.

2 Material and Methods

This section analyses the state-of-the-art regarding the different subsystems previously presented: smart grid control, video-surveillance, and TSN.

2.1 Smart Grid Control and Monitoring

An RTU connects devices in the physical world to a Supervisory Control And Data Acquisition (SCADA) system by transmitting telemetry data to a master system, and by using messages from the master supervisory system to control connected devices. A variety of protocols is used to communicate with RTUs [16]. For this specific implementation, an RTU counts with a control unit and an acquisition module. The head control unit of the RTU performs the control functions for the complete system, centralizes the information acquired by other modules, executes programmable logic operations, and manages the communication protocols and the specific user applications. The acquisition module of the RTU performs the monitoring and control of the electrical equipment, transformers, or sensors.

High-availability Seamless Redundancy (HSR) is an Ethernet network protocol standardized by the IEC 62439-3 [17]. It allows zero-time recovery in a single failure, and the redundancy in the communication is transparent to the application level.

2.2 Smart Video Surveillance

One of the main challenges faced by security systems is the inability to automatically operate without the supervision of a human operator [18]. As a result, there has recently been a growing interest in smart video surveillance systems that automatically detect certain events such as intrusions [19], abandoned objects [20] or fires [21] without requiring constant human observation [22]. Humans cannot operate around the clock, they need to sleep and even when they are awake they do not remain fully focused for long [18].

There are also examples in the state of the art of monitoring systems for work environments. For example, with high-resolution cameras, the increased memory and processing capacities, and greater Internet availability, new systems have been proposed for monitoring construction areas in transport projects [23]. In such a way, safety management has been radically improved in these challenging environments. In the case of smart grid infrastructures and substations in particular, security and safety are a fundamental requirement due to the high voltage equipment. Thus, it is required to efficiently detect and track workers, intruders or pedestrians near the substation based on video surveillance to ensure people and infrastructure safety [24]. Among the leading techniques in recent years, the most prominent include the use of Convolutional Neural Networks (CNN) [25].

2.3 Time Sensitive Networks

The TSN solution implements the key standards to deliver deterministic QoS (bounded latency with low jitter and guaranteed bandwidth) for time-critical traffic in the presence of best-effort traffic. Different data streams are differentiated and prioritized by means of IEEE 802.1Q VLAN tagging. Attending to the specified VLAN priority, data streams are queued and forwarded following the time-aware traffic shaper scheme defined on the IEEE 802.1Qbv, based on a strict time-driven cyclic schedule. The generalized Precision Time Protocol (gPTP, IEEE 802.1AS) enables the stringent coordination between network elements and time-critical distributed applications to provide the required end-to-end bandwidth and latency guarantees. Despite the novelty of TSN, this technology is being applied to different fields such as the aerospace [26] and on automotive [27] industries, industrial automation [14] or Industrial Internet of the Things (IIoT) applications [28].

3 Proposed Approach

This work proposes a heterogeneous CPS that combines the electrical substation control and the video surveillance subsystems using the TSN network. Both services operate using

the same network and without interfering with each other. In this way, low latency transmission of critical substation control traffic is guaranteed in the presence of bandwidth-intensive video transmission. Fig. 1 illustrates an overview of the deployment proposing three different levels:

- The **smart grid control subsystem** is in charge of the control and monitoring of the substation equipment (via the IEC 60870-5-104 protocol and SAS [29, 30]). It is also responsible for the monitoring of alarms triggered by the surveillance subsystem (via Modbus/TCP protocol) if a restricted perimeter violation is detected. This information is used to discharge the electrical substation and to change the operating mode of all RTUs installed in the electrical substation with the aim of guaranteeing the security and safety of the critical infrastructures and the intruder. The operating modes are:
 - **Remote:** The control of the equipment is remotely operated from a SCADA system [31]
 - **Local:** The operator is allowed to take control and avoid the need to interact or accept remote supervisory commands. One should take into account that *local* in this context always refers to the locality of the actual equipment. This mode avoids safety risks for intruders in the electrical substation [31]

Additionally, this subsystem ensures the reliability through HSR.

- The **smart video surveillance subsystem** detects and tracks workers or intruders, or monitors secured perimeters or protection zones within the substation. An intrusion detection causes an alarm that leads to sending Modbus/TCP commands to the smart grid control subsystem. Computation in this subsystem is distributed between a cloud server and networked SoC nodes: video surveillance tasks such as person detection are carried out in the local edge nodes, while the cloud server is used for people tracking and facility perimeter control.
- The **TSN** guarantees the coexistence of the two previous subsystems sharing the same network: for SAS monitoring and substation control tasks, it provides low latency and low deviation. For smart surveillance control messages, TSN guarantees necessary bandwidth usage. Finally, TSN bounds the latency of SAS alarm network traffic triggered by the surveillance subsystem.

3.1 Smart Grid Control Subsystem

The smart grid subsystem ensures the reliability, monitors and controls the SAS, particularly enabling the local and remote modes. It also manages alarms triggered by the video surveillance subsystem. When an intruder is detected,

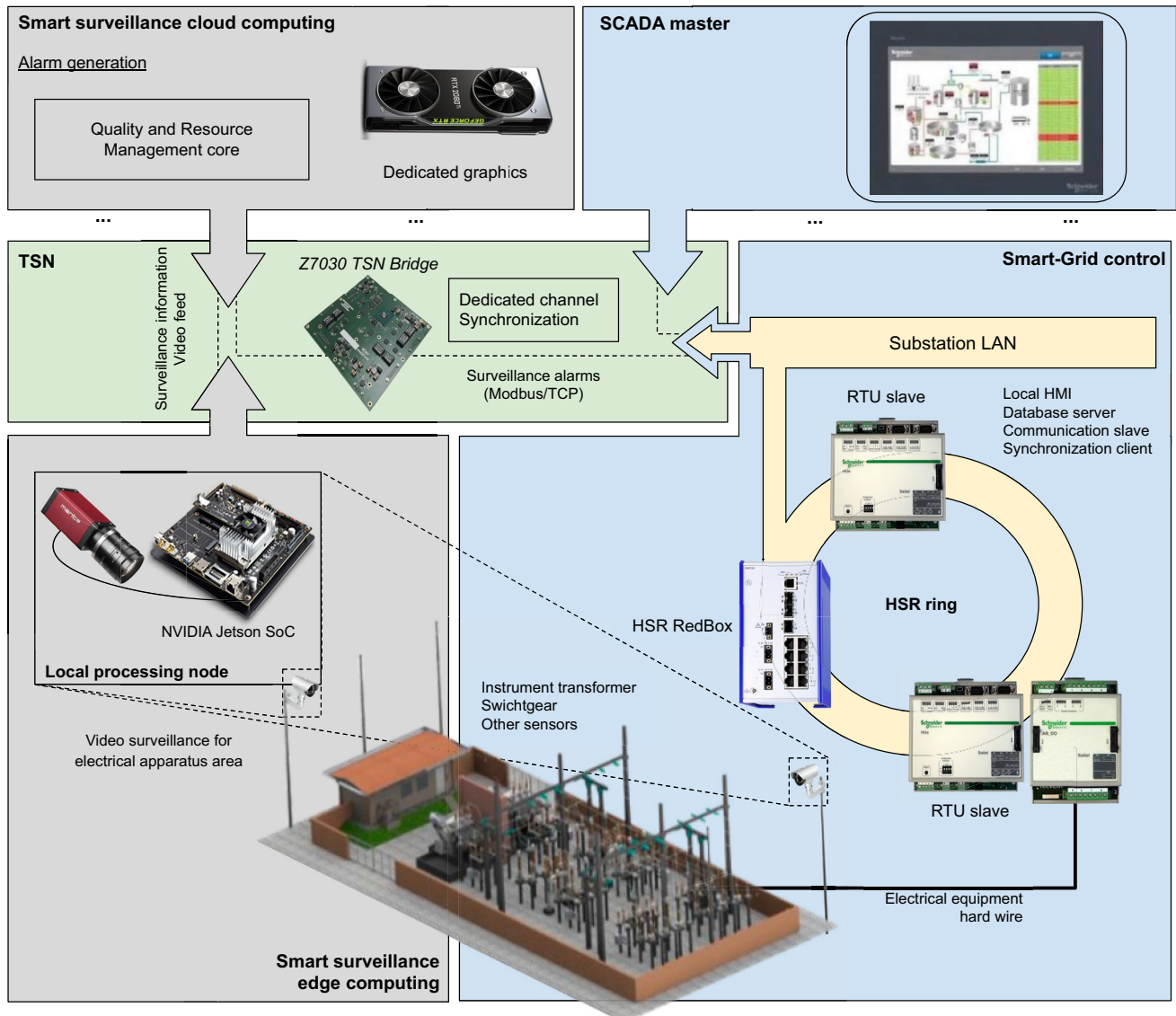


Figure 1 Overview of our CPS for protecting smart grid critical infrastructures: At the control level, RTUs are responsible for the monitoring and control of the electrical substation infrastructure and implementing the HSR ring to withstand single communication failures. Regarding the smart surveillance subsystem, the distributed video processing is done in embedded high-performance System-on-Chip (SoCs) nodes [15]. They perform human detection via Deep

Learning (DL). The cloud server carries out the management of multiple edge data to track people and monitor the substation limits and desired perimeters. Simultaneously, the cloud server also determines the optimal edge configuration based on information from the Quality and Resource Management (QRM) system. All of the aforementioned devices are connected via the TSN network by means of *Z7030 TSN Bridges*.

a discharge for the electrical substation and a redistribution of its load around the smart grid is managed by the SCADA system with the RTUs using IEC 60870-5-104 protocol, in order to:

- Prevent the electricity supply to be affected if there is any damage to the physical infrastructure on the electrical substation by the intruder
- Guarantee the safety of the intruder if any element with high electrical risk is touched

Then a command is received by all RTUs disabling the remote mode, to reject any remote command that may be executed in the electrical substation that could harm the intruder, and blocking local mode to ensure that the intruder does not execute commands within the substation (see the complete work flow in Fig. 2).

In order to ensure that all RTUs receive the previously mentioned command to disable the remote mode when an intrusion is detected (even during single point communication failure situations), HSR protocol is used. In this case,

the deployed topology is an HSR ring in which command messages are duplicated to guarantee that e.g. alarms from the SCADA to perform the substation discharge are reliably transmitted to all RTUs. Simultaneously, HSR nodes continuously check for duplicates to avoid performing an action twice.

3.2 Smart Surveillance Subsystem

Smart video surveillance tasks are distributed in multiple processing platforms: at the edge, local nodes connected to the surveillance cameras perform image acquisition and preprocessing and detection of persons, extracting their location and their appearance features. The cloud server, on the other hand, first gathers all results from edges. Then, it tracks locally detected people within the substation facility and monitors the perimeters under supervision. Finally, the cloud server is also responsible for the communication with the smart grid subsystem when alarms are triggered.

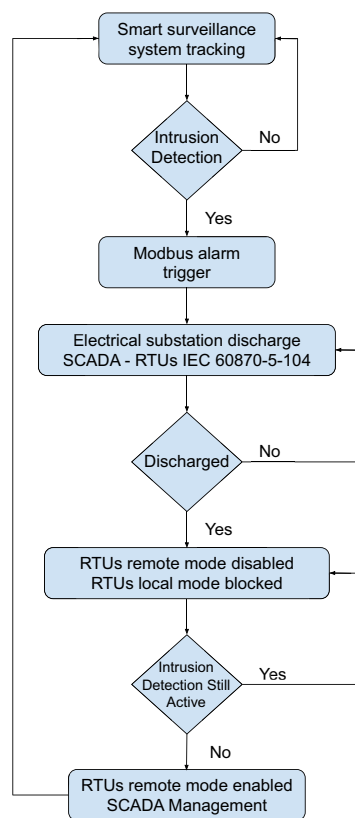


Figure 2 Flow chart for the local/remote management modes on the smart grid control subsystem, processing alarms generated by the surveillance subsystem. When an intrusion detection is triggered by the smart surveillance subsystem, the smart grid control subsystem manages the discharge of the affected substation, if the intrusion is still active the remote mode is disabled and the local mode is blocked. Only when the intrusion alarm ceases, the remote mode is enabled again.

3.2.1 Camera Calibration and Adjustment of Regions of Interest

When placing a video camera within the facility to be video monitored, the camera tilt, pan and height must be taken into consideration in order to optimize the captured area of the scene. If the camera is placed higher with little tilt, it points to a smaller area of the ground than if it is placed lower, pointing parallel to the ground plane. The latter camera configuration covers more ground plane area and consequently, captures areas that might not be interesting for video surveillance tasks.

Thus, in order to define the region of the image that is relevant for each camera, a preprocessing stage to define the area of interest of their field of view (FoV) is performed. This operation consists of estimating the area on the scene ground plane that are able to contain up to 2-meter height targets, filtering out targets that are far away or partially outside the image. The process of this calculation and its result for one of the deployed cameras is shown in Fig. 3. For this height estimation, it is necessary to take the horizon line as a reference. Thus, when the surveillance subsystem deployed cameras are calibrated [32], the homographic transformation matrix between the camera perspective plane in pixels and the aerial map view with cartographic coordinates is applied. Then, having the cartographic longitude and latitude lines in the camera scene, the horizon line is estimated, namely where these lines converge. By taking the height of an object in the image as a reference and projecting it on the horizon, the measurement of any other part of the image can be found.

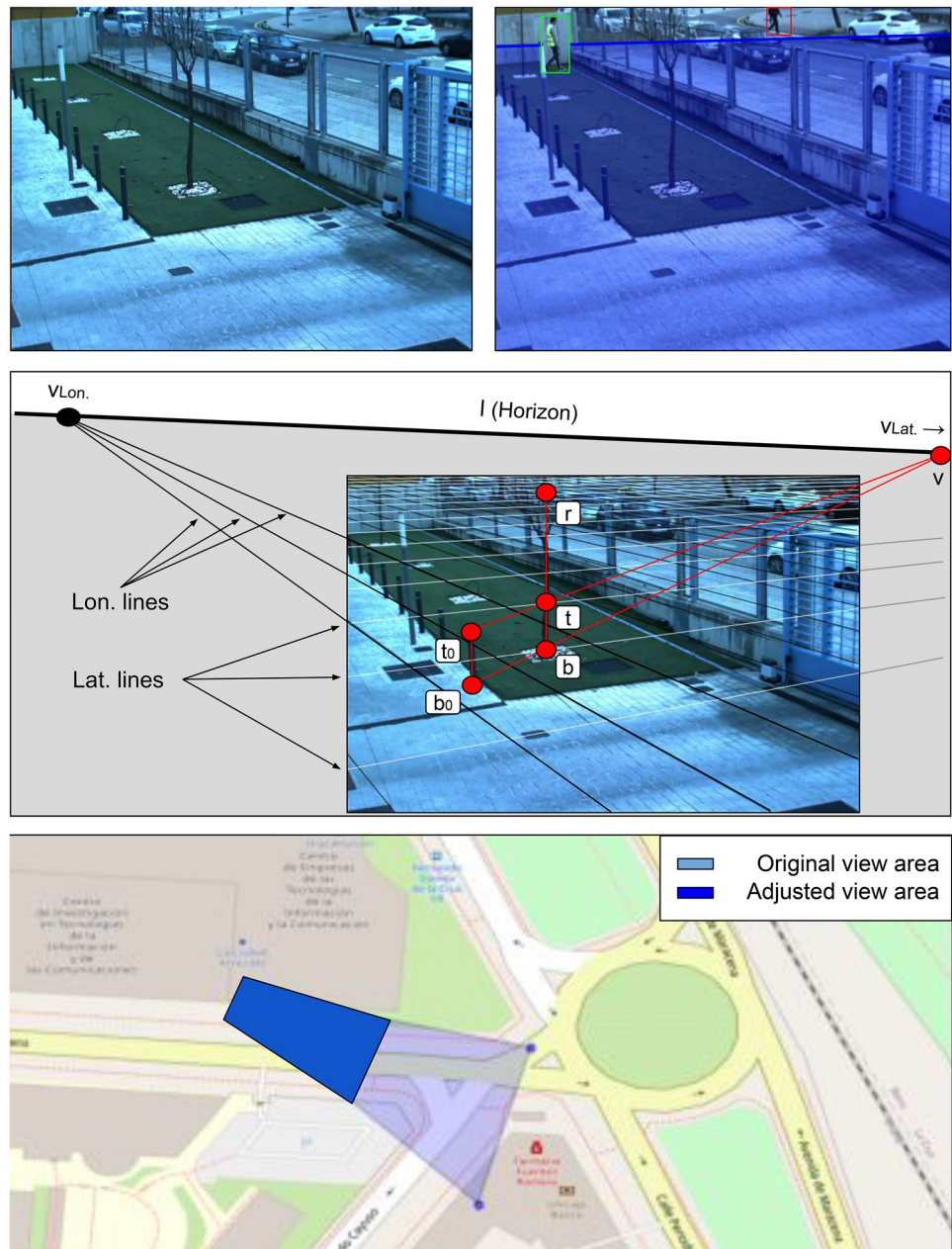
In line with the nomenclature shown in the middle picture of Fig. 3, we get V_{lon} and V_{lat} where the longitude and latitude lines converge respectively and l is obtained from connecting this two vanishing points. Vanishing point v is formed by connecting b_0 and b points with a line and intersecting it with the l line. And t is the projection of t_0 point on the reference object, using v point as centre of projection. Taking into consideration the aforementioned values, the target height is estimated as in Eq. (1), where h is the reference object height and h' is the target height to be estimated.

$$\frac{\|t - b\| * \|v_z - r\|}{\|r - b\| * \|v_z - t\|} = \frac{h}{h'} \quad (1)$$

3.2.2 Edge Processing

Our CPS for critical infrastructure protection is a heterogeneous distributed system whose edge nodes are NVIDIA Jetson TX2 and NVIDIA Jetson Xavier SoCs [15]. On these local nodes, video acquisition from the cameras and video processing for initial surveillance tasks are carried

Figure 3 Calculation of ROIs for surveillance cameras, (from top to bottom and left to right): 1) View from a surveillance camera. 2) Result after adjustment, the ground plane is considered as the ROI where a full-body person could be detected (up to 2 meters of height). The frame shows the detection of a person (in green) inside the ROI and one missed person outside this ROI and partially outside the image boundaries (in red). 3) Adjustment process shown graphically: $Lon.$ and $Lat.$ are the longitude and latitude cartographic lines respectively; V_{lon} and V_{lat} are the longitude and latitude vanishing points; l is the horizon line; b is the base point and r the higher point or height of the reference object; b_0 is the base point and t_0 the higher point of the object to measure; t is the projection of t_0 in the reference object and v is the vanishing point of the object to measure. 4) Comparison on map between the original (light blue) and the adjusted (dark blue) view areas.



out. These compute-intensive tasks require GPU hardware acceleration to be performed in real time. This makes the NVIDIA Jetson TX2 and NVIDIA Jetson Xavier SoCs very good candidates since they are equipped with GPUs with 256 and 512 CUDA cores [15] respectively. Next, we detail the surveillance tasks performed by each local node.

Peterson Detection The aim is to locate parts of the image that correspond to a person. Initially, a background subtraction method is used to differentiate the objects in the moving foreground from the rest of the scene, which is known as background and remains static. The background subtraction

method is based on Mixture of Gaussians (MOG) model [33]. As the name suggests, this method mixes multiple weighted Gaussian distributions to model each pixel in the background of an image. The mixture weights represent the proportions of time these colors remain in the scene. The most likely background colors are those that remain the longest and most static in the scene. In contrast, in the foreground area potential targets will be located.

Locations in the foreground are our Regions of Interest (ROIs), where people are detected using a classification technique based on Machine Learning. Through a Deep Convolutional Neural Network (DCNN) model [32, 34],

trained with large datasets of images showing people [35, 36], each ROI is assigned a probability to contain a person. This DCNN model is based on the MobileNetV2 network architecture [34] and optimised for embedded devices through TensorRT SDK [37], which reduces the size and complexity of the model. A ROI is marked as a person if the DCNN model assigns it a likelihood $p \geq 0.85$.

Deep Feature Extraction In order to reidentify people detected in previous frames, it is necessary to define the different unique identities. By estimating a feature identifying vector, the main physical visual attributes based on appearance are stored in a summarized container for later comparison. This feature vector is a numeric array in which each element matches a different area of the ROI marked as a person. A specific combination of values is unique to a person's image and it is enough for a successful identification. In order to get this feature vector, a DCNN model [38] calculates through aggregation and weighting operations the importance of the different areas of the image of a person such as: pixel RGB values, edges, patterns found at a higher level, etc. After applying this step, the original ROI marked as a person is reduced to a 128-length float vector. This DCNN model has also been optimized for better performance on embedded devices using the TensorRT SDK.

The optimization of the previous models, both detection and deep feature extraction, using TensorRT SDK, has resulted in a more efficient inference engine for each case. To achieve this, the number of parameters of the original models is reduced by merging layers and tensors. It also makes for more efficient memory management, allowing a better parallel execution of DCNN models. In addition, the complexity of the model weight type is reduced (from 32-bit float to 8-bit integer). In the optimization of both inference tasks, the overall model QoS has been significantly increased: performance has been increased by 24% while model fit metrics have been reduced by less than 0.01%.

3.2.3 Cloud Processing

The cloud server provides greater computing power and dedicated GPU for improving processing performance. With this, the scalability of the system is guaranteed with local nodes on the edge processing and transmitting a large amount of data. Next, we detail the surveillance tasks performed by the cloud server.

Multi-person Tracking Using the detection information and the person feature descriptor collected from the local nodes, people are tracked around the whole substation [38]. The tracking method looks for matches of the new detection values with those of a tracks database, that are continuously updated. The similarities between detection features as appearance and

location and the same features of the tracks are used to create a distance matrix. The assignment to the correct track is solved using the classic Hungarian algorithm [39].

Perimeter Monitoring By knowing the real-time location of each person tracked inside the substation it is possible to trigger alarms when secured perimeters or protection zones are violated. A perimeter database is available on the cloud server. Each database entry stores the points (*lat*, *lon*) that form the polygon area contained within the perimeter.

Communication with the Control Subsystem When a person is detected or an intruder breaks into a secured perimeter an alarm is directly sent to the substation control RTU via the MODBUS/TCP protocol. The cloud server acts as the master in the connection, reading the RTU status and registering the alarms in its memory.

3.2.4 FIVIS Monitoring Dashboard

Data processing and visualization is an important aspect of CPS, specially regarding monitoring of complex critical infrastructures with multiple subsystems. *FIVIS* is an extension of the *IVIS-CORE* framework [40] that supports storage, analysis, and visualization of monitoring data. This tool makes it possible to run custom analysis on data from multiple sources, whose results are used as input information for dashboards (see Fig. 4) and specialized reports or formatted data streams for other machines. In this CPS there are data from multiple sources and at different levels:

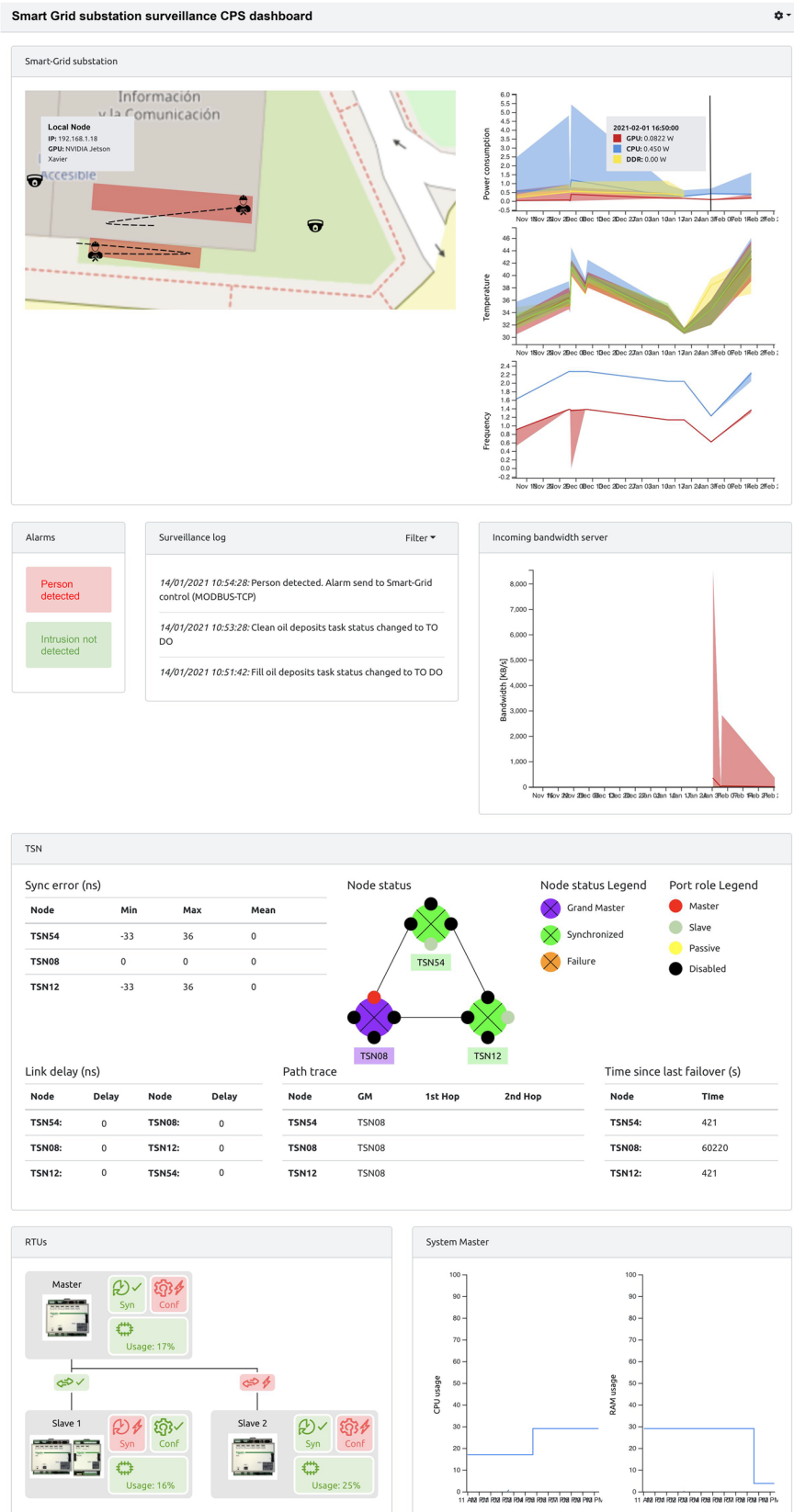
- **Hardware statistics:** Status and resource utilization of the different platforms of each of the subsystems of the CPS.
- **TSN network status:** Synchronization between network nodes and port status of each node.
- **Surveillance information:** Location and trajectory of workers/intruders within the substation, perimeter security status, and surveillance alarm notification.

3.3 TSN for Smart Grid

In our system, the TSN component delivers deterministic bounded latency and guarantees bandwidth for time-critical traffic, while in the presence of best-effort video traffic from multiple cameras that consumes most of the data bandwidth. In our CPS, three TSN bridges have been deployed creating a ring topology. Each TSN bridge is based on a Zynq-7000 FPGA, supporting four 100/1000-Base-T interfaces. Each interface VLAN classifies the different traffic types on transmission and route ingressing VLAN-tagged data streams.

Firstly, in our case, the correct synchronization between the components is ensured via gPTP. Regarding reliability, it is enabled through the IEEE 802.1AS standard that

Figure 4 Final CPS dashboard: Monitoring information is displayed for each of the subsystems (top: video surveillance; middle: TSN; bottom: smart grid monitoring and control), as well as hardware information for the different platforms. As observed, depending on the type and nature of the information to be represented, a different kind of representation is used: the evolution of numerical data over time is represented in line charts or scatter plots; data whose relevance is related to the exact time when they are generated are displayed as a discrete amount; data already processed are shown in a text log.



defines adaptation mechanisms to face network failures. Besides the link status provided by the 1000-Base-T physical layer, the link propagation delay is continuously monitored to assure, on one hand, accurate recovering of the remote time reference and, on the other hand, support of hard real-time communication. The interface is considered faulty if the link propagation delay reaches a threshold or if the remote peer cannot cooperate on the measurement. If unreachable, the grand Master or network time reference is re-elected, and the role of the interfaces are adapted to receive (Slave) or re-transmit (Master) the synchronization information. Synchronization may be received from different paths, in case of redundant network topologies. A network interface may behave as passive, to back the slave interface in case of failure.

Secondly, this implementation considers up to four different traffic priorities. Each output interface manages up to four VLAN-priority queues: time-synchronization messages, time-critical commands, control traffic, and best effort. The time-aware traffic shaper has been implemented on forwarding to provide deterministic QoS for the highest priorities, and isolation against best-effort data streams. Network time synchronization is required for the stringent coordination between time-triggered Smart Grid distributed nodes and the TSN bridges participating along the transmission path. Additionally, smart-grid MODBUS commands and smart-surveillance detection streams are subject to bandwidth guarantee QoS. Finally, the best-effort traffic is used for video streaming.

4 Results

All the data collected in the CPS is aggregated and displayed in a dashboard within the *FIVIS* platform (see Section 3.2.4). This dashboard displays both surveillance and alarm information (top), as well as the status of the substation control platforms (bottom), and the status of the TSN network and its nodes (middle). This dashboard, shown in Fig. 4, collects the following information from each of the subsystems:

- **Surveillance:** Node, perimeter and person locations within the substation; power consumption, temperature, frequency and bandwidth usage for the different local node hardware components; alarm status (person detection and intrusion); text log for relevant events (e.g. alarms, broken perimeters, or communication messages with the substation control RTUs).
- **TSN status:** Synchronization status and delay of each network node, as well as delays between the nodes. Time since last network failover.

- **Substation control RTUs:** Connection and synchronization status of the HSR ring RTU components; configuration status, RAM, and CPU usage of each RTU device.

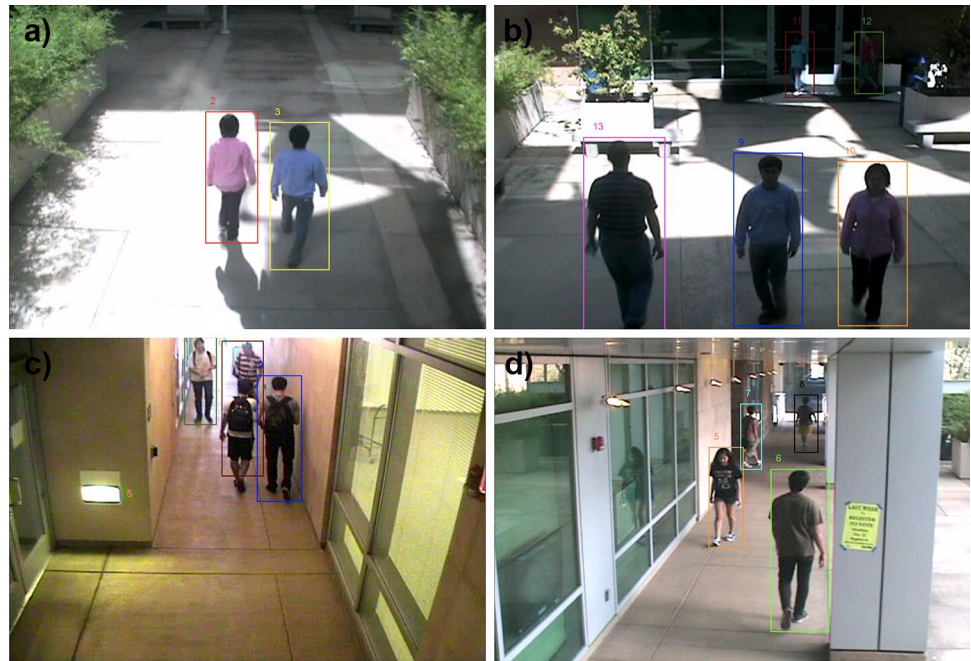
4.1 Human Detection and Tracking Evaluation

As already discussed, the main smart video surveillance tasks performed in our system are detection and tracking of multiple people. Both tasks are performed consecutively and, in order to compare their quality with state-of-art methods, these tasks are evaluated with the Camera Network Tracking Dataset (CamNeT) [41]. Besides, to compare the results of our people detector + tracker with this dataset, the MOT Challenge metrics are used. Those metrics are common in the Multiple Object Tracking (MOT) problem and include [42]:

- **Multi-object tracking accuracy (MOTA):** overall tracking accuracy in terms of false positives, false negatives and identity switches
- **Multi-object tracking precision (MOTP):** overall tracking precision in terms of bounding box overlap between ground-truth and reported location
- **Mostly tracked (MT):** percentage of ground-truth tracks that have the same label for at least 80% of their life span
- **Mostly lost (ML):** percentage of ground-truth tracks that are tracked for at most 20% of their life span
- **Identity switches (ID):** number of times the reported identity of a ground-truth track changes

The results of our detection + tracking method are shown in Table 1 and compared with two state-of-the-art methods. *FaitMOT* [44] proposes a multi-branch DCNN that integrates detection and tracking tasks together by reidentifying the detected subjects in subsequent frames. *Tracktor* [43] addresses the tracking of multiple people by adapting DCNN-based people detectors to include a regression layer that allows the temporal readjustment of new people locations detected in successive frames. While the two methods compared do not consider performance as a critical aspect that defines their QoS, ours proposes an architecture to be embedded in edge devices. In this line, the other methods achieve results close to real time exclusively using dedicated, high-performance graphic processing (dGPU). The results are similar to the other methods, considering that our solution works in real-time (> 30 fps) and in a distributed manner. Also, our method is embedded in local edge devices, adapted and optimized to deal with limited resources. Fig. 5 shows some qualitative results for our method on the dataset used for testing. The inferior quality of our system that is shown on the metrics is partially justified due to the optimization to low resource devices. In addition, as mentioned previously in Section 3.2.1, our system does not consider

Figure 5 Detection + MOT results for the CamNeT dataset [41]. **a**, **b**, and **d** show the correct track results for people in the scene, including unique identifiers and bounding boxes; **c** shows a detection fail of a partially occluded person, resulting in a detection bounding box containing two people.



people who are partially out of the scene, as they are not fully detectable. However, this is not the case for the other two comparative methods. As a consequence, part of the track path is omitted when they are leaving or coming into the scene and as a result there is, for example, a higher number of false negatives.

4.2 HSR Redundancy Frames

As mentioned before, control and monitoring traffic within the electric substation is critical. No packages are to be missed and high-availability and fast recovery are required for smart grid communication. In this section, the smart grid subsystem is implemented using an HSR ring that ensures that at any time, two different paths reach a single node. In this way, the redundant path delivers a copy of any original command message. Finally, for the normal operation of the electric substation, duplicates are filtered out.

Figure 6 shows two real duplicate frame messages on the HSR ring of the smart grid control subsystem, ensuring the redundancy in a single communication network failure. The smart grid control communications are properly encapsulated on two identical frames with the same sequence

number and sent through two different paths of the HSR ring to prove the reliability of the network. The two messages are identical except for the origin (different paths and lane ids).

4.3 Demonstration Scenario in the Event of an Alarm

Figure 7 shows an example of the CPS operation, where a person is detected within the substation limits and the subsequent alarm is triggered. This alarm causes the substation to switch to an operating mode and discharge in order to protect the well-being of the person and the electric substation equipment.

In detail, on the left side a camera view of the substation is shown that is processed to remove the background (foreground in white and background in black). Next, the foreground ROIs are passed to a classifier that detects if a person is included within them. In the example, both ROIs are analyzed but only the one on the top, that contains a person in a yellow vest with his back turned (likely an operator) is selected. The second ROI contains a moving car that is therefore discarded. The CNN-based model extracts the person descriptor (feature vector) that is sent along with its

Table 1 MOT results for CamNeT dataset.

	MOTA ↑	MOTP ↑	MT ↑	ML ↓	FN ↓	FP ↓	ID ↓
Ours	86.9%	88.2%	84.8%	10.2%	7.2%	4.3%	6.9%
Tracktor [43]	88.1%	91.6%	86.3%	7.5%	6.4%	3.2%	5.4%
FairMOT [44]	91.9%	93.4%	90.1%	5.7%	5.5%	1.9%	2.8%

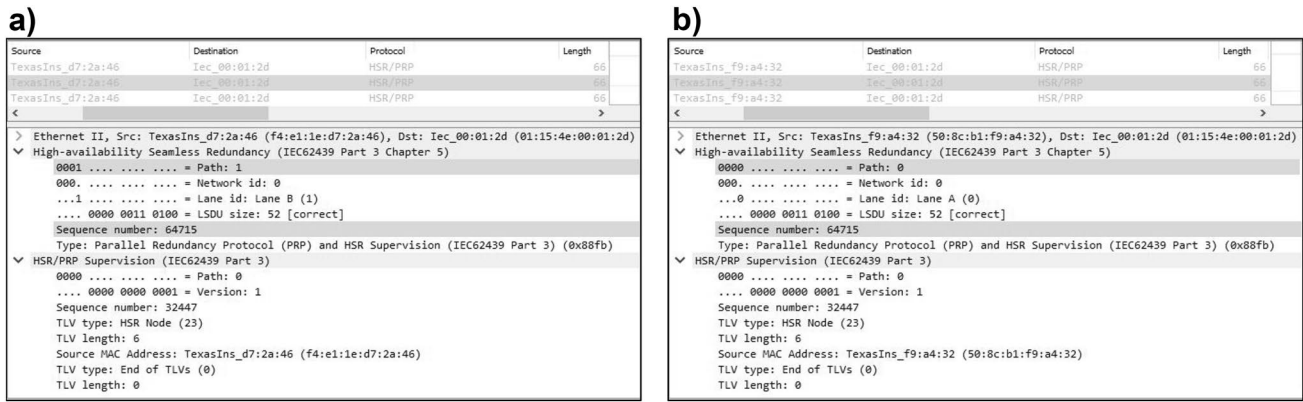


Figure 6 Traffic sniffed on eth0 (a) and eth1 (b), two network interfaces of an HSR node. Captures prove the redundant messages that ensure the reliability of the critical communication: both traces are

identical, having the same sequence number except that they have been captured from two different paths/lanes of the HSR ring.

location (pixel coordinates) to the cloud server. This information is then compared with the tracks in the database at the cloud server. If the detection feature vector or its location is similar to one of the saved tracks, the track is automatically updated. In the example, the worker attributes are already in this database and thus, they are only updated. Otherwise, a new track would be added and initialized. With the track location up-to-date, it is verified that it does cross

the limits of one of the perimeters. Therefore, a MODBUS command is sent to the SAS RTU master as an alarm. The MODBUS frame includes some interesting fields: the *MB slave* field indicates the slave id we are communicating with (slave 01 in the example); the *Function* field specifies the type of operation performed (01 - read coils); and the result of that read with the alarm value is reported in the *Alarm value* field (01 - Perimeter violated, in the example).

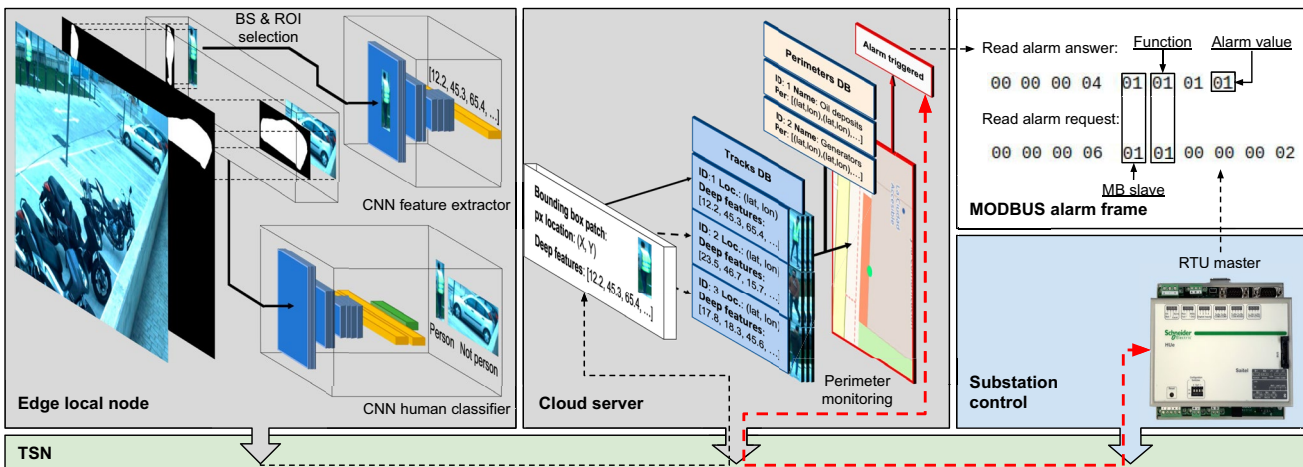


Figure 7 Example scenario of an alarm triggered when a person violates the electric station perimeter. (Left Edge local nodes process video from surveillance cameras, detecting areas with moving objects (ROI) that are selected only if they contain a person. The next step calculates a feature vector based on the person appearance and sends it to the cloud server. In the example, a moving car is also selected as a ROI but then discarded when classified as a *not a person*. Center)

The cloud server stores all positive detections and person trajectory tracks in a database (*Tracks DB*), grouping all the same person detections/tracks and determining if the person is breaking a secured perimeter or protection zone. Right) If this is the case, the appropriate alarm is triggered and read by the master HSR RTU using MODBUS/TCP protocol.

5 Conclusions

Surveillance is essential for the safety of substations and personnel or intruders. The proposed CPS guarantees the security of the electricity supply and the safety of people in the facility. In our case, this is demonstrated through the automatic management of smart video surveillance that generates alarms, the electric control that acts performing substation discharges and switching remote/local mode of the SAS.

The coexistence of the different subsystems ensures that: IEC 60870-5-104 traffic from the smart grid subsystem is not affected, corrupted, lost or delayed due to the integration of the three subsystems. MODBUS communications between the smart video surveillance and the smart grid subsystems are possible, delivering alarms to the smart grid subsystem to act accordingly. Finally, the integration of HSR allows communications even in situations with a single communication failure.

This integrated video surveillance and substation control CPS improves the security of the smart grid critical infrastructures. Smart video surveillance, whose tasks have been optimised for the processing platforms used, makes it possible to provide real-time alarms for dangerous situations. Furthermore, by using a TSN, the communication of these alarms with the SAS is guaranteed. This work opens the door to the use of heterogeneous subsystems within the same network, while facilitating intercommunication between in advance, independent systems without compromising the operation of any of them.

Acknowledgements This work was partially supported by the EU Project FitOptiVis [45] through the ECSEL Joint Undertaking under GA n. 783162, a Spanish National grant funded by MINECO through APCIN PCI2018-093184. The work was also partially supported by the National Grant PID 2019-109434RA-I00/SRA (State Research Agency 10.13039/501100011033). Gabriel Jimenez-Perera's research has been supported by "Consejería de Transformación Económica, Industria, Conocimiento y Universidades de la Junta de Andalucía" (Grant Ref. PREDOC_00280).

Funding Funding for open access publishing: Universidad de Granada/CBUA.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Commission, E. (2020). *on the EU Security Union Strategy*. Brussels: European Commission.
2. Jiangtao, H. (2019). Discussion on The Construction of Substation Security Video Surveillance System. *IOP Conference Series: Materials Science and Engineering*, 563, 032004.
3. Orfanidis, G., Apostolidis, S., Kapoutsis, A., Ioannidis, K., Kosmatopoulos, E., Vrochidis, S., & Kompatsiaris, I. (2019). Autonomous Swarm of Heterogeneous Robots for Surveillance Operations. In *Lecture Notes in Computer Science* (pp. 787–796). Springer volume 11754 LNCS.
4. Ahmed, C. M., & Kandasamy, N. K. (2021). A comprehensive dataset from a smart grid testbed for machine learning based cps security research. *Cyber-Physical Security for Critical Infrastructures Protection*, 12618, 123.
5. Adamiak, M., & Kulshrestha, A. (2002). *Design and implementation of a UCA based substation control system*. Lett: Phys. Rev.
6. Vahidinasab, V., Tabarzadi, M., Arasteh, H., Alizadeh, M. I., Mohammad Beigi, M., Sheikhzadeh, H. R., Mehran, K., & Sepasian, M. S. (2020). Overview of Electric Energy Distribution Networks Expansion Planning. *IEEE Access*.
7. Kumar, S., Das, N., & Islam, S. (2015). High performance communication redundancy in a digital substation based on IEC 62439-3 with a station bus configuration. In *2015 Australasian Universities Power Engineering Conference: Challenges for Future Grids, AUPEC 2015*.
8. Deniz, D., Barranco, F., Isern, J., & Ros, E. (2020, September). Reconfigurable cyber-physical system for lifestyle video-monitoring via deep learning. In *2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, 1, 1705–1712. IEEE. <https://doi.org/10.1109/ETFA46521.2020.9211910>
9. Li, J., Wang, H., Zhao, Y., Huang, R., & Yang, S. (2019). Application Research of Artificial Intelligent Technology in Substation Inspection Tour. In *2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)* Itaic (pp. 1097–1100). IEEE.
10. Gauci, A., & Schneider, P. (2012). Smart Grid Fault Location, Isolation, and Service Restoration (FLISR) Solutions to Manage Operational and Capital Expenditures.
11. Wang, T., Wang, K., Li, J., Yu, H., Shuai, W., Bian, J., & Zhao, X. (2017). Fast recognition of human climbing fences in transformer substations. In *2017 Ninth International Conference on Advanced Computational Intelligence (ICACI)* (pp. 195–200). IEEE.
12. Sau, C., Rinaldi, C., Pomante, L., Palumbo, F., Valente, G., Fanni, T., ... & de Beer, A. (2021). Design and management of image processing pipelines within CPS: Acquired experience towards the end of the FitOptiVis ECSEL Project. *Microprocessors and Microsystems*, 87, 104350. <https://doi.org/10.1016/j.micpro.2021.104350>
13. Slater, J., Nesbitt, A., Morison, G., & Boreham, P. (2019). A Hybrid Cloud for Data Analytics in Electrical Substation Condition Monitoring Systems. In *2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)* (pp. 283–288). IEEE.
14. Lo Bello, L., & Steiner, W. (2019). A Perspective on IEEE Time-Sensitive Networking for Industrial Communication and Automation Systems. *Proceedings of the IEEE*, 107, 1094–1120.
15. NVIDIA. Embedded Systems for Next-Generation Autonomous Machines. NVIDIA Jetson: The AI platform for autonomous everything.
16. Electric, S. (2016). IEC Smart Energy Roadmap; v3.0e.

17. International Electrotechnical Commission. (2018). Industrial communication networks. High availability automation networks. Parallel Redundancy Protocol (PRP) and High-availability Seamless Redundancy (HSR). IEC 62439-3:2018 International Electrotechnical Commission.
18. Alshammari, A., & Rawat, D. B. (2019). Intelligent Multi-Camera Video Surveillance System for Smart City Applications. In *2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)* (pp. 0317–0323). IEEE.
19. Kim, S. H., Lim, S. C., & Kim, D. Y. (2018). Intelligent intrusion detection system featuring a virtual fence, active intruder detection, classification, tracking, and action recognition. *Annals of Nuclear Energy*, *112*, 845–855.
20. Luna, E., Miguel, J. C. S., Ortego, D., & Martínez, J. M. (2018). Abandoned object detection in video-surveillance: Survey and comparison. *Sensors (Switzerland)*, *19*.
21. Roque, G., & Padilla, V. S. (2020). LPWAN Based IoT Surveillance System for Outdoor Fire Detection. *IEEE Access*.
22. Park, H., Park, S., & Joo, Y. (2019). Robust detection of abandoned object for smart video surveillance in illumination changes. *Sensors (Switzerland)*, *19*.
23. Luo, H., Liu, J., Fang, W., Love, P. E., Yu, Q., & Lu, Z. (2020). Real-time smart video surveillance to manage safety: A case study of a transport mega-project. *Advanced Engineering Informatics*, *45*, 101100.
24. Peng, Q., Luo, W., Hong, G., Feng, M., Xia, Y., Yu, L., Hao, X., Wang, X., & Li, M. (2016). Pedestrian detection for transformer substation based on Gaussian mixture model and YOLO. *Proceedings - 2016 8th International Conference on Intelligent Human-Machine Systems and Cybernetics, IHMSC 2016*, *2*, 562–565.
25. Wang, H., Zhang, X., Sun, Y., Li, J., & Li, Y. (2019). Research and application of artificial technology for substation environment surveillance system. In *Proceedings of 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference, ITAIC 2019*.
26. Steiner, W., Heise, P., & Schneelee, S. (2014). Recent IEEE 802 developments and their relevance for the avionics industry. In *2014 IEEE/AIAA 33rd Digital Avionics Systems Conference (DASC)* (pp. 2A2-1–2A2-12). IEEE.
27. Kim, H. J., Choi, M. H., Kim, M. H., & Lee, S. (2021). Development of an ethernet-based heuristic time-sensitive networking scheduling algorithm for real-time in-vehicle data transmission. *Electronics (Switzerland)*, *10*, 1–11.
28. Metaal, M. A., Guillaume, R., Steinmetz, R., & Rizk, A. (2020). Integrated Industrial Ethernet Networks: Time-sensitive Networking over SDN Infrastructure for mixed Applications. *IFIP Networking 2020 Conference and Workshops, Networking 2020*, (pp. 803–808).
29. Lin, C.-Y., & Nadjm-Tehrani, S. (2018). Understanding IEC-60870-5-104 Traffic Patterns in SCADA Networks. In *Proceedings of the 4th ACM Workshop on Cyber-Physical System Security* (pp. 51–60). New York, NY, USA:ACM
30. Radoglou-Grammatikis, P., Sarigiannidis, P., Giannoulakis, I., Kafetzakis, E., & Panaousis, E. (2019). Attacking IEC-60870-5-104 SCADA Systems. In *2019 IEEE World Congress on Services (SERVICES)* (pp. 41–46). IEEE.
31. Smith, C. (2013). Remote Control Modes - Local vs Remote.
32. Isern, J., Barranco, F., Deniz, D., Lesonen, J., Hannuksela, J., & Carrillo, R. R. (2020). Reconfigurable cyber-physical system for critical infrastructure protection in smart cities via smart video-surveillance. *Pattern Recognition Letters*, *140*, 303–309.
33. Zivkovic, Z. (2004). Improved adaptive Gaussian mixture model for background subtraction. In *International Conference on Pattern Recognition* (pp. 28–31). volume 2.
34. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (pp. 4510–4520).
35. Harvey Adam. LaPlace, J. (2019). MegaPixels: Origins, Ethics, and Privacy Implications of Publicly Available Face Recognition Image Datasets.
36. Oh, S., Hoogs, A., Perera, A., Cuntoor, N., Chen, C.-C., Lee, J. T., Mukherjee, S., Aggarwal, J. K., Lee, H., Davis, L., Swears, E., Wang, X., Ji, Q., Reddy, K., Shah, M., Vondrick, C., Pirsivash, H., Ramanan, D., Yuen, J., Torralba, A., Song, B., Fong, A., Roy-Chowdhury, A., & Desai, M. (2011). A large-scale benchmark dataset for event recognition in surveillance video. In *CVPR 2011* (pp. 3153–3160). IEEE.
37. NVIDIA (2020). Tensorrt. <https://developer.nvidia.com/tensorrt>.
38. Wojke, N., Bewley, A., & Paulus, D. (2017). Simple online and realtime tracking with a deep association metric. In *ICIP* (pp. 3645–3649). IEEE.
39. Kuhn, H. W. (1956). Variants of the hungarian method for assignment problems. *Naval Research Logistics Quarterly*, *3*, 253–258.
40. Bulej, L., Bures, T., Hnetyka, P., Camra, V., Siegl, P., & Topfer, M. (2020). IVIS: Highly customizable framework for visualization and processing of IoT data. In *2020 46th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)* (pp. 585–588). IEEE.
41. Zhang, S., Staudt, E., Faltemier, T., & Roy-Chowdhury, A. K. (2015). A Camera Network Tracking (CamNeT) Dataset and Performance Baseline. In *2015 IEEE Winter Conference on Applications of Computer Vision* (pp. 365–372). IEEE.
42. Bernardin, K., & Stiefelhagen, R. (2008). Evaluating multiple object tracking performance: The Clear MOT metrics. *EURASIP Journal on Image and Video Processing*, *2008*, 1–10.
43. Bergmann, P., Meinhardt, T., & Leal-Taixé, L. (2019). Tracking without bells and whistles. In *The IEEE International Conference on Computer Vision (ICCV)*.
44. Zhang, Y., Wang, C., Wang, X., Zeng, W., & Liu, W. (2020). Fairmot: On the fairness of detection and re-identification in multiple object tracking. arXiv preprint <http://arxiv.org/abs/2004.01888>.
45. Al-Ars, Z., et al. (2019) The fitoptivis eysel project: Highly efficient distributed embedded image/video processing in cyber-physical systems. In: Proceedings of the 16th ACM International Conference on Computing Frontiers, CF '19, p. 333338. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3310273.3323437>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Juan Isern PhD student in Advanced Monitoring Techniques for Distributed Environments and High Performance Vision Processing Architectures at the University of Granada, Spain, where he also got his MSc. in Data Science and Computer Engineering. He has worked as a Machine Learning Engineer in the field of Smart Cities and in public capital projects such as ECSEL H2020 FITOPTIVIS.



Gabriel Jimenez-Perera is with the Department of Computer Engineering, Automatics and Robotics at the University of Granada, Spain. He received the M.Sc. degree in Data Science and Computer Engineering in 2018 and is currently a PhD Candidate. He works on federated learning, focusing on computer vision and embedded systems.



Luis Medina-Valdes received his Ph.D. in Electronics in 2018 from the University of Alcalá (Spain), after a research grant on the Spanish Research Council (CSIC). From 2018 to 2022, he worked as R&D technical manager at Orolia Spain. Previously, he was a researcher at the University Institute of Computer Research (Alicante, Spain) working on FPGA-based algorithms for

autonomous vehicles. He has been participating and managing several industrial aerospace and R&D European funded projects. His major interests are ultra-precise time synchronisation and deterministic communication applied to mission critical infrastructures.



Pablo Chaves Telecom Engineer from the University of Granada, graduated in 2015. He finished, in 2018, an Official Master's Degree in Managing Renewable Energy Installations and Internationalization of Projects from the University of Cardenal Herrera. In 2014 he joined Schneider Electric working in Substation Automation RTU validation and certification and in R&D collaborative funded projects such as ECSEL H2020 FITOPTIVIS, H2020 IELECTRIX,

H2020 SABINA, FP7 IDE4L and national projects such as 3S-CS, SAGRA and Ferrosmartgrid. In 2021 he started to work on Micro Grid line of business inside Schneider Electric, promoting solutions for the automation and optimization of the microgrids management.



David Pampliega received the Telecom Engineering degree from the Industrial Technical Engineering School of Seville, Spain, 2008. He is currently enrolled in a PhD program in the University of Seville. He is Senior Project Manager in the Digital Power Line of Business of Schneider Electric, with more than 10 years of experience in managing Smart Grid R&D projects in European programs, involving

different utilities all over Europe. He has a technical background in different smart grid areas, including the monitoring of MV/LV substations, RTU, microgrids, cybersecurity, smart meters or data analytics among others. In addition, he is responsible for driving customer interaction with respect to advanced technical queries related to grid digital controllers.



Francisco Ramos PhD, M.S in Electrical Engineering, in Computer & Network Engineering and in Physics and Executive MBA. He started working in Alcatel in Brussels and then moved to Switzerland joining EMMicroelectronics-Swatch Group where he worked for 7 years. From 2005 he is working in Schneider Electric where was appointed as Innovation Manager in charge of R&D collaborative projects in programs such HE, H2020, ECSEL among others. Manage more than 100 projects and 19 Million Euro funding. 18 published papers, 5 patents. From May 2022 appointed as PowerLogic T300 Substation Controller Product Owner, managing about 40 Million Euro Business within Distribution Automation segment.



Francisco Barranco received his M.Sc. degree in Computer and Network Engineering, and his Ph.D. degree in Computer Engineering from the University of Granada, Spain, in 2008 and 2012, respectively. Currently, he is an Associate Professor with the Department of Computer Architecture and Technology, University of Granada. His research interests include robotics, embedded real-time machine vision, bio-inspired processing, and cognitive vision.