**Journal of Signal Processing Systems manuscript No.**
(will be inserted by the editor)

# A Cyber-Physical System for integrated remote control and protection of smart grid critical infrastructures

**Juan Isern*** · **Gabriel Jimenez-Perera** · **Luis Medina-Valdes** · **Pablo Chaves** · **David Pampliega** · **Francisco Ramos** · **Francisco Barranco**

Received: date / Accepted: date

**Abstract** This work proposes a Cyber-Physical System (CPS) for protecting Smart Electric Grid Critical Infrastructures (CI) using video surveillance while remotely monitoring them. Due to the critical nature of 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 (TSN) 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 (QoS), latency and bandwidth guarantees, required by the time-critical Smart Grid control. On the one hand, the CPS utilizes High-availability Seamless Redundancy (HSR) in the control subsystem via Remote Terminal Units (RTU) 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 CI. Moreover, it is also able to directly interoperate with RTUs via MODBUS protocol to send alarms in case of e.g. 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 · High-availability Seamless Redundancy · Remote Terminal Unit

*corresponding author: jisern@ugr.es

Juan Isern · Gabriel Jimenez-Perera · Francisco Barranco
Computer Architecture and Technology CITIC,
University of Granada, Granada 18014, Spain
Tel.: +34-958-241-775
E-mail: jisern@ugr.es
E-mail: gabrieljimenez@ugr.es
E-mail: fbarranco@ugr.es

Luis Medina-Valdes
Seven Solutions,
Granada 18014, Spain
E-mail: luis.medina@sevensols.com

Pablo Chaves · David Pampliega · Francisco Ramos
Schneider Electric España S.A.,
Sevilla 41092, Spain
E-mail: pablo.chaves@se.com
E-mail: david.pampliega@se.com
E-mail: francisco.ramos@se.com

## 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 disruption (physical or digital) [8].

Over the last decades, the requirements for the stability and reliability of the power supply have increased with the continuous development of our society [14]. The 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 the citizens and the industry. Such control of CI is a task of great complexity, as a consequence of the several risks that the security personnel has to supervise. It has been shown that investing on preventive equipment and security personnel resources is often insufficient. Thus, autonomous surveillance systems that improve the protection and reduce cost are needed [29]. Preferably, in-

tegrated systems that enable seamless interoperation across the subsystems in electrical substations and allow for joint strategies are the aim of future solutions [2].

The main elements of the electrical distribution network are the substation automation systems (SAS) [1] which monitor and control the electrical infrastructure. Distribution networks comprise from distribution substations to the service entrance of the the electricity consumers, including distribution substations, primary feeders, distribution transformers, and secondary systems [38]. 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 voltages levels or to switch between alternating and direct current. A key element of the SAS are the 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 contribute to the reliability by decreasing probability of communication failures [18].

While intelligent video analytics is the most widely used technology globally in security, the use of this technology is not widely deployed in electric substations in general, and in final distribution substations in particular [19]. 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 [40,14]. However, this continuous on-line monitoring produces vast volumes of surveillance data that must be analyzed in a timely and efficient manner, while avoiding the interference with the electric control and monitoring.

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 [35].

The second part is achieved by establishing data traffics 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 CI. Time Sensitive Networking is a set of extensions of the IEEE 802.1 and 802.3 standards adding deterministic QoS to bridged Ethernet networks, such as bounded latency and guaranteed bandwidth [22]. 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 CI 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 CI, 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 analysis 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, controlled by a microprocessor, connects devices in the physical world to a distributed control system or 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 are used to communicate with RTUs [9]. 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 mon-

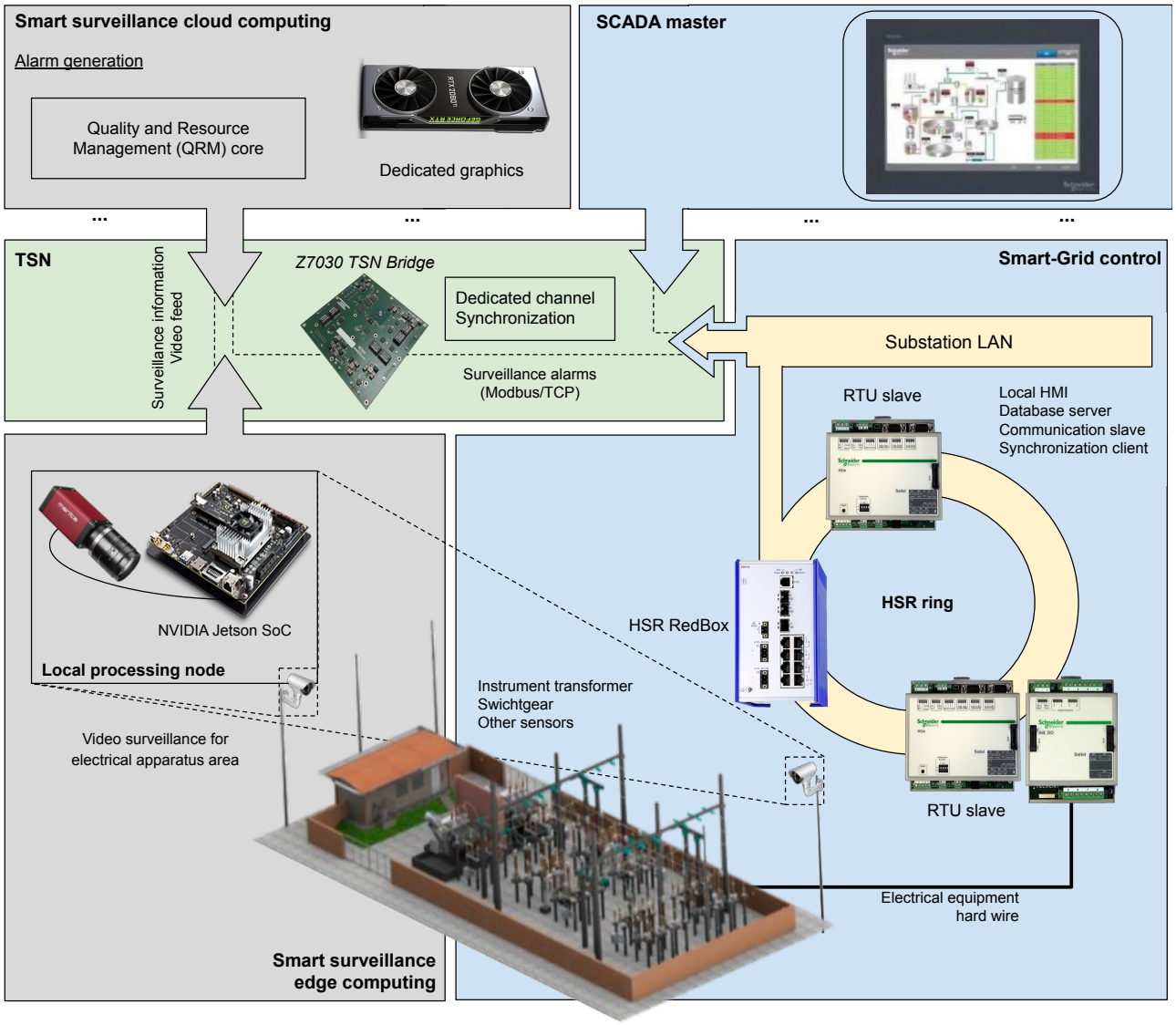


**Fig. 1** Overview of our Smart Grid critical infrastructure protection CPS: At the control level, RTUs are responsible for the monitoring and control of the electrical substation infrastructure and implementing the HSR ring to withstand single communication failures. Regarding the smart surveillance subsystem, the distributed local edge nodes process the video in embedded high-performance SoCs (System-on-Chip Nvidia Jetson devices [26]). 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 QRM system. All of the aforementioned devices are connected via the TSN network by means of *Z7030 TSN Bridges*.

itoring and control of the electrical equipment, transformers, or sensors.

HSR is an Ethernet network protocol standardized by the IEC 62439-3 [12]. It allows zero-time recovery in a single failure, and the redundancy 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 [4]. As a result, there has recently been a growing interest in smart video surveillance systems that automatically detect certain events such as intrusions [16], abandoned objects [23] or fires [33] without requiring constant human observation [30]. Humans cannot operate around the clock, they need to

sleep and even when they are awake they do not remain undistracted for long [4].

With high-resolution video cameras, increased memory and processing capacities and greater Internet availability, the ability to report construction and maintenance of an asset has been radically improved [24]. In the case of Smart Grid infrastructures and substations in particular, security is a fundamental requirement due to the high voltage. 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 [31]. Among the leading techniques in recent years, the most prominent include the use of convolutional neural networks (CNN) [39].

### 2.3 Time Sensitive Networks

The Time Sensitive Network (TSN) solution implements the key standards to deliver deterministic QoS (bounded latency with low jitter and guaranteed bandwidth) for time-critical traffics 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 [37] and on automotive [15] industries, industrial automation [21] or IIoT applications [25].

## 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 [32, 20]. It is also responsible for the monitoring of alarms triggered by the surveillance subsystem (via Modbus/TCP) 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 CI and the intruder. The operating modes are:
  - **Remote**: The control of the equipment is remotely operated from a SCADA (supervisory control and data acquisition) system [36]
  - **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 [36]

  Additionally, this subsystem ensures the reliability through High-availability Seamless Redundancy (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 (System on Chip) 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. Specifically, when an intruder is detected in a restricted perimeter of the electrical substation an alarm is received on the Smart Grid control subsystem from the Smart surveillance subsystem via the MODBUS protocol. At this very moment, a discharge for the electrical substation

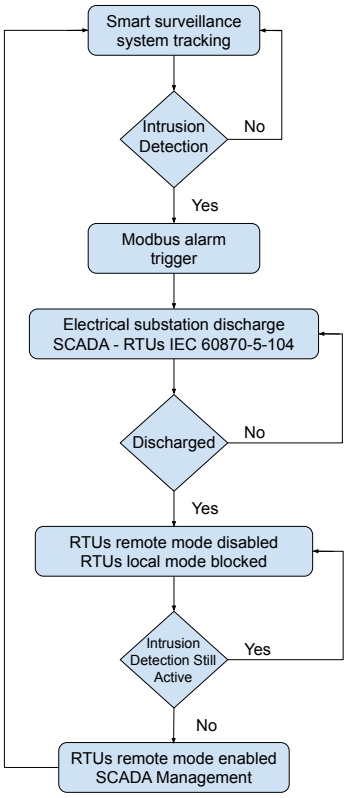


**Fig. 2** Flow chart for the local/remote modes management on the Smart Grid control subsystem, processing alarms generated by the Smart surveillance subsystem. When an intrusion detection is triggered by the Smart surveillance subsystem, the Smart Grid control subsystem manages the discharge of the 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.

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 the intruder cannot execute commands within the substation. 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), a High-availability Seamless Redundancy (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, firstly 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.

### *3.2.1 Camera calibration and surveillance ROI adjustment*

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 [13], 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 refer-

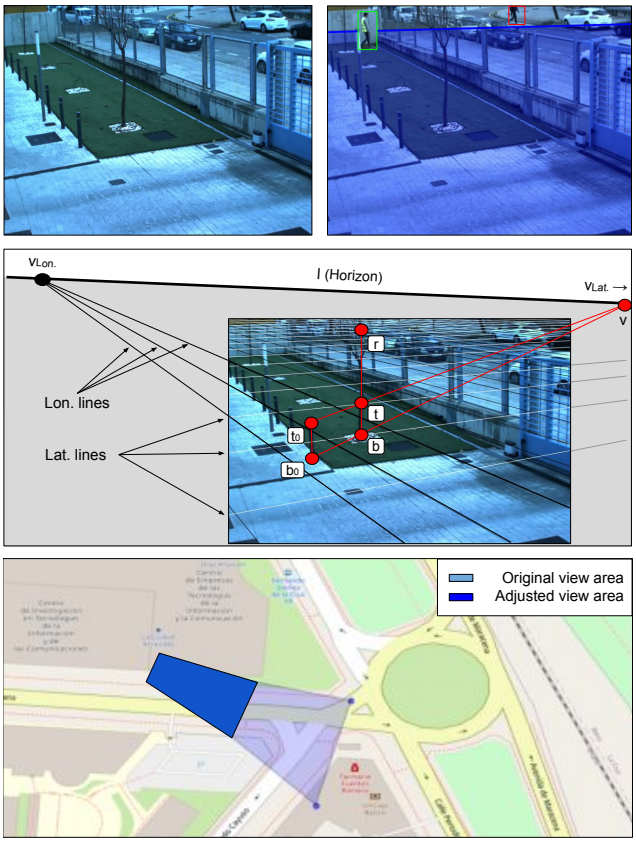


**Fig. 3** Process for calculating ROIs (Region of Interest) for surveillance cameras. From top to bottom and from left to right: 1) View from one of the surveillance cameras. 2) Result after adjustment, showing the ground plane area considered as the ROI, where a full-body person is detected (up to 2 meters height). The frame shows the successful detection of a person (in green) inside the region of interest and a missed person detection that is outside the region and partially outside the image boundaries (in red). 3) Adjustment process shown graphically. Where: *Lon.* lines and *Lat.* lines are 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 camera viewing area (light blue) and adjusted area (dark blue).

ence 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.

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

### *3.2.2 Edge processing*

As mentioned before, our proposed CPS for CI protection is a heterogeneous distributed system whose edge nodes are NVIDIA Jetson TX2 and NVIDIA Jetson Xavier SoCs [26]. On these local nodes, video acquisition from the cameras and video processing for initial surveillance tasks are carried 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 [26] respectively. The following are details of the surveillance tasks performed at each local node:

- **Person detection**: The objective is to find and define the parts of the image that corresponds 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 [44]. 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, the foreground is the area where potential targets will be located.
  Locations in the foreground are our regions of interest (ROI), where people are detected using a classification technique based on Machine Learning. Through a deep convolutional neural network (DCNN) model [13,34], trained with large datasets of images showing people [28,11], the each ROI is assigned a probability to contain a person. This DCNN model is based on Google's MobileNetV2 [34] structure and optimised for embedded devices through TensorRT SDK [27], 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 [41] 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.

#### *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. The cloud server takes care of the tasks listed below:

- **Multi-person tracking**: Using the detection information and the person feature descriptor collected from the local nodes, people are tracked around the whole substation [41]. 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 [17].
- **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 substation 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 registers and writing 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 [7] that supports storage, analysis, and visualization of monitoring data. This tool makes it possible to run custom analyses 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 Time Sensitive Network component delivers deterministic bounded latency and guarantees bandwidth for time-critical traffics, 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 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 dif-

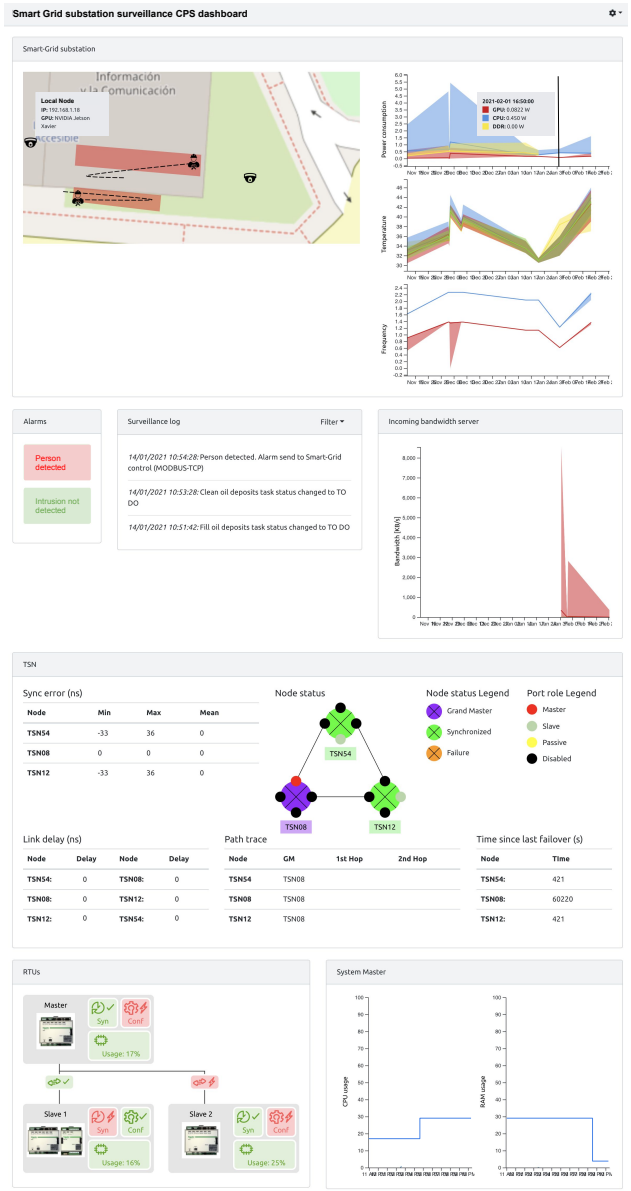


**Fig. 4** Final CPS dashboard: Monitoring information is displayed for each of the subsystems (top: video surveillance; middle: TSN; bottom: smart grid monitoring and control), as well as hardware information for the different platforms. As can be observed, depending on the type and nature of the information to be represented, a different kind of representation is used: Numerical data that require their evolution over time are represented in line charts or scatter plots. Data whose importance is limited to the moment it is generated is displayed as a KPI with its own value. Data that is relevant because it has already been processed is shown in a text log.

ferent 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. Attending this our system application traffics, 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. Besides, Smartgrid Modbus commands and smart-surveillance detection streams are object of 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 Figure 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) [42]. 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 [6]:

- **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

**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 [5] | 88.1% | 91.6% | 86.3% | 7.5% | 6.4% | 3.2% | 5.4% |
| FairMOT [43] | 91.9% | 93.4% | 90.1% | 5.7% | 5.5% | 1.9% | 2.8% |

The results of our detection + tracking method are shown in Table 1 and compared with two state-of-the-art methods (Tracktor [5] and FairMOT [43]). 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 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's 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.

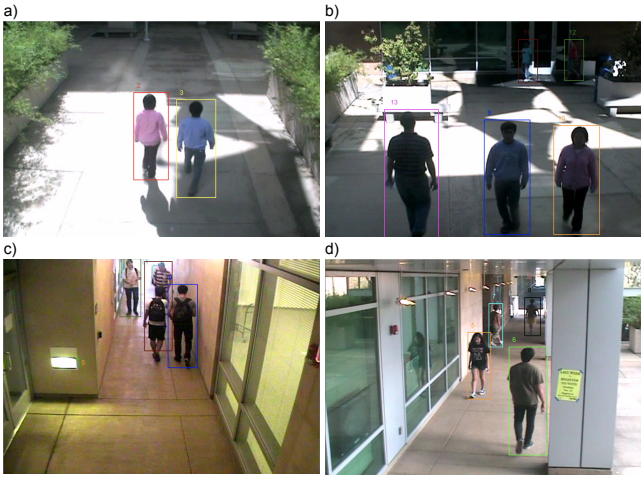


**Fig. 5** Detection + MOT results for the CamNeT dataset [42]. a, b and d show the correct track results for people in scene, including unique identifiers and bounding boxes; c shows a detection fail of a partially occluded persons, resulting in a detection bounding box containing two people.

Fig. 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

Fig. 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 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

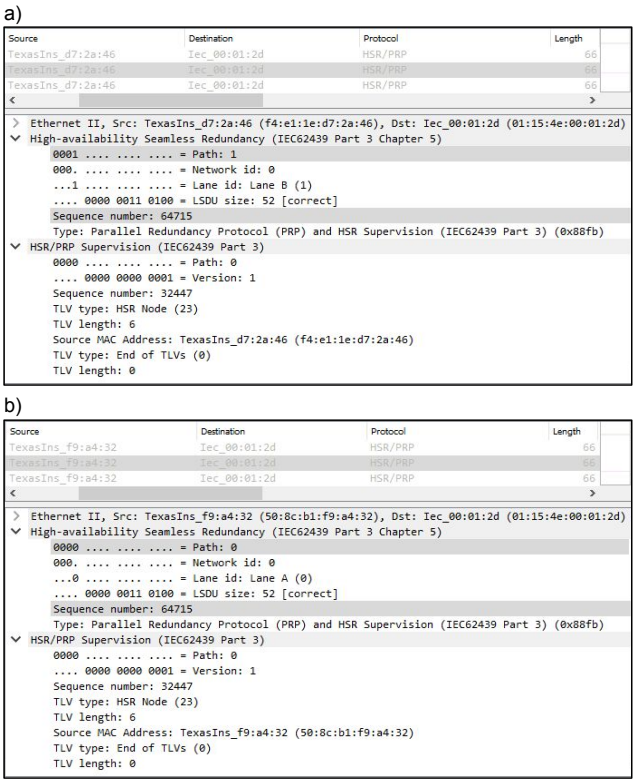


**Fig. 6** Traffic sniffed on eth0 (a) and eth1 (b), two network interfaces of an HSR node. Captures prove the redundant messages that ensure the reliability of the critical communication: both traces are identical, having the same sequence number except that they have been captured from two different paths/lanes of the HSR ring.

similar to one of the saved tracks, the track is automatically updated. In the example, the worker attributes are already in this database and are updated. If this was not the case, a new track is added and initialized with such information. Anyway, with the track location up-to-date, it is verified that it does not exceed the limits of any of the perimeters stored in the database. In this situation, as in the example, 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 *Alarm* value field (01 - Perimeter violated, in the example).

## 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 CI. Smart video surveillance, whose tasks have been optimised for the processing platforms used, makes it possible to provide real-time danger alarms. 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 [3] through the ECSEL Joint Undertaking under GA n. 783162, a Spanish National grant funded by MINECO through APCIN PCI2018-093184, and partially by the Research Network RED2018-102511-T.

## References

1. Adamiak, M., Kulshrestha, A.: Design and implementation of a UCA based substation control system. Phys. Rev. Lett (2002)
2. Ahmed, C.M., Kandasamy, N.K.: A comprehensive dataset from a smart grid testbed for machine learning based cps security research. Cyber-Physical Security for Critical Infrastructures Protection **12618**, 123 (2021)
3. Al-Ars, Z., et al.: The fitoptivis ecsel 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. 333–338. Association for Computing Machinery, New York, NY, USA (2019). DOI 10.1145/3310273.3323437
4. Alshammari, A., Rawat, D.B.: 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 (2019). DOI 10.1109/CCWC.2019.8666579. URL https://ieeexplore.ieee.org/document/8666579/
5. Bergmann, P., Meinhardt, T., Leal-Taixé, L.: Tracking without bells and whistles. In: The IEEE International Conference on Computer Vision (ICCV) (2019)

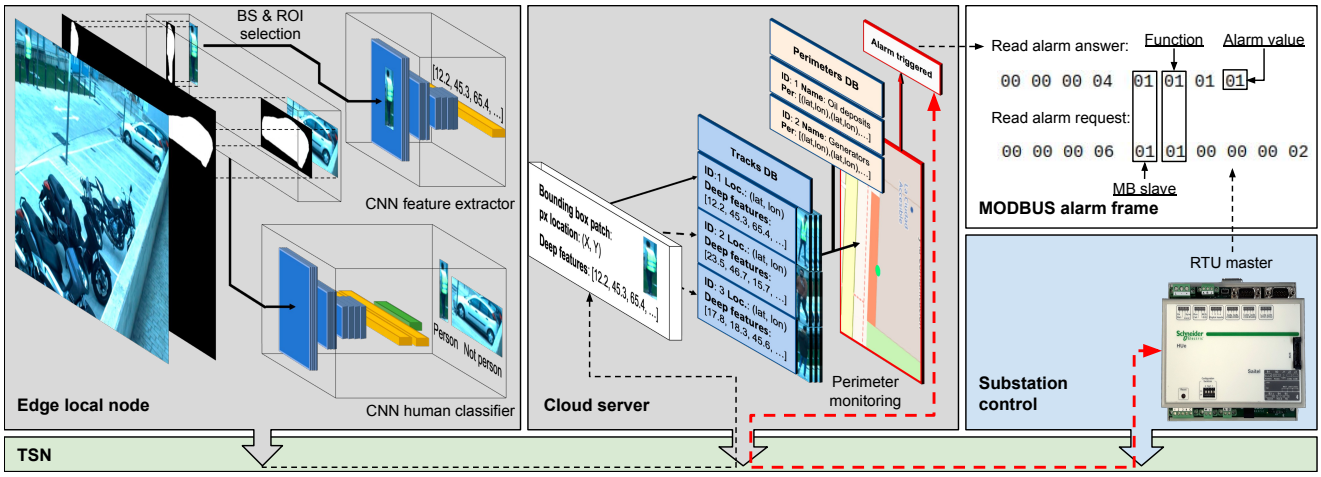


**Fig. 7** Example scenario of an alarm triggered when a person violates the electric station perimeter. Left) Edge local nodes process video from surveillance cameras, detecting areas with moving objects (ROI) that are selected only if they contain a person. The next step calculates a feature vector 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


6. Bernardin, K., Stiefelhagen, R.: Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics. EURASIP Journal on Image and Video Processing **2008**, 1–10 (2008)
7. Bulej, L., Bures, T., Hnetynka, P., Camra, V., Siegl, P., Topfer, M.: 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 (2020). DOI 10.1109/SEAA51224.2020.00095. URL https://ieeexplore.ieee.org/document/9226306/
8. Commission, E.: on the EU Security Union Strategy. European Commission,, Brussels (2020). URL https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0605
9. Electric, S.: IEC Smart Energy Roadmap; v3.0e (2016). URL https://www.se.com/uk/en/product-subcategory/80262-remote-terminal-unit/
10. Gauci, A., Schneider, P.: Smart Grid Fault Location , Isolation , and Service Restoration ( FLISR ) Solutions to Manage Operational and Capital Expenditures (2012)
11. Harvey Adam. LaPlace, J.: MegaPixels: Origins, Ethics, and Privacy Implications of Publicly Available Face Recognition Image Datasets (2019). URL https://megapixels.cc/
12. International Electrotechnical Commission: Industrial communication networks. High availability automation networks. Parallel Redundancy Protocol (PRP) and High-availability Seamless Redundancy (HSR). Standard IEC 62439-3:2018, International Electrotechnical Commission (2018). URL https://webstore.iec.ch/publication/24447
13. Isern, J., Barranco, F., Deniz, D., Lesonen, J., Hannuksela, J., Carrillo, R.R.: Reconfigurable cyber-physical system for critical infrastructure protection in smart cities via smart video-surveillance. Pattern Recognition Letters **140**, 303–309 (2020). DOI 10.1016/j.patrec.2020.11.004. URL https://linkinghub.elsevier.com/retrieve/pii/S0167865520304098
14. Jiangtao, H.: Discussion on The Construction of Substation Security Video Surveillance System. IOP Conference Series: Materials Science and Engineering **563**, 032004 (2019). DOI 10.1088/1757-899X/563/3/032004
15. Kim, H.J., Choi, M.H., Kim, M.H., Lee, S.: Development of an ethernet-based heuristic time-sensitive networking scheduling algorithm for real-time in-vehicle data transmission. Electronics (Switzerland) **10**(2), 1–11 (2021). DOI 10.3390/electronics10020157
16. Kim, S.H., Lim, S.C., Kim, D.Y.: Intelligent intrusion detection system featuring a virtual fence, active intruder detection, classification, tracking, and action recognition. Annals of Nuclear Energy **112**, 845–855 (2018). DOI 10.1016/j.anucene.2017.11.026. URL https://doi.org/10.1016/j.anucene.2017.11.026
17. Kuhn, H.W.: Variants of the hungarian method for assignment problems. Naval Research Logistics Quarterly **3**(4), 253–258 (1956). DOI 10.1002/nav.3800030404. URL http://doi.wiley.com/10.1002/nav.3800030404
18. Kumar, S., Das, N., Islam, S.: 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 (2015). DOI 10.1109/AUPEC.2015.7324838
19. Li, J., Wang, H., Zhao, Y., Huang, R., Yang, S.: 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 (2019). DOI 10.1109/ITAIC.2019.8785585. URL https://ieeexplore.ieee.org/document/8785585/
20. Lin, C.Y., Nadjm-Tehrani, S.: 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. ACM, New York, NY, USA (2018). DOI 10.1145/3198458.3198460. URL https://dl.acm.org/doi/10.1145/3198458.3198460
21. Lo Bello, L., Steiner, W.: A Perspective on IEEE Time-Sensitive Networking for Industrial Communication and

Automation Systems. Proceedings of the IEEE **107**(6), 1094–1120 (2019). DOI 10.1109/JPROC.2019.2905334
22. Lo Bello, L., Steiner, W.: A perspective on ieee time-sensitive networking for industrial communication and automation systems. Proceedings of the IEEE **107**(6), 1094–1120 (2019). DOI 10.1109/JPROC.2019.2905334
23. Luna, E., Miguel, J.C.S., Ortego, D., Martínez, J.M.: Abandoned object detection in video-surveillance: Survey and comparison. Sensors (Switzerland) (2018). DOI 10.3390/s18124290
24. Luo, H., Liu, J., Fang, W., Love, P.E., Yu, Q., Lu, Z.: Real-time smart video surveillance to manage safety: A case study of a transport mega-project. Advanced Engineering Informatics **45**(December 2019), 101100 (2020). DOI 10.1016/j.aei.2020.101100. URL https://doi.org/10.1016/j.aei.2020.101100
25. Metaal, M.A., Guillaume, R., Steinmetz, R., Rizk, A.: Integrated Industrial Ethernet Networks: Time-sensitive Networking over SDN Infrastructure for mixed Applications. IFIP Networking 2020 Conference and Workshops, Networking 2020 pp. 803–808 (2020)
26. NVIDIA: Embedded Systems for Next-Generation Autonomous Machines. NVIDIA Jetson: The AI platform for autonomous everything. URL https://www.nvidia.com/en-us/autonomous-machines/embedded-systems/
27. NVIDIA: Tensorrt. https://developer.nvidia.com/tensorrt (2020)
28. 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., Pirsiavash, H., Ramanan, D., Yuen, J., Torralba, A., Song, B., Fong, A., Roy-Chowdhury, A., Desai, M.: A large-scale benchmark dataset for event recognition in surveillance video. In: CVPR 2011, pp. 3153–3160. IEEE (2011). DOI 10.1109/CVPR.2011.5995586. URL http://ieeexplore.ieee.org/document/5995586/
29. Orfanidis, G., Apostolidis, S., Kapoutsis, A., Ioannidis, K., Kosmatopoulos, E., Vrochidis, S., Kompatsiaris, I.: Autonomous Swarm of Heterogeneous Robots for Surveillance Operations. In: Lecture Notes in Computer Science, vol. 11754 LNCS, pp. 787–796. Springer (2019)
30. Park, H., Park, S., Joo, Y.: Robust detection of abandoned object for smart video surveillance in illumination changes. Sensors (Switzerland) **19**(23) (2019). DOI 10.3390/s19235114
31. Peng, Q., Luo, W., Hong, G., Feng, M., Xia, Y., Yu, L., Hao, X., Wang, X., Li, M.: 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 (2016). DOI 10.1109/IHMSC.2016.130
32. Radoglou-Grammatikis, P., Sarigiannidis, P., Giannoulakis, I., Kafetzakis, E., Panaousis, E.: Attacking IEC-60870-5-104 SCADA Systems. In: 2019 IEEE World Congress on Services (SERVICES), pp. 41–46. IEEE (2019). DOI 10.1109/SERVICES.2019.00022. URL https://ieeexplore.ieee.org/document/8817093/
33. Roque, G., Padilla, V.S.: LPWAN Based IoT Surveillance System for Outdoor Fire Detection. IEEE Access (2020). DOI 10.1109/ACCESS.2020.3003848
34. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C.: MobileNetV2: Inverted Residuals and Linear Bottlenecks. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition pp. 4510–4520 (2018). DOI 10.1109/CVPR.2018.00474
35. Slater, J., Nesbitt, A., Morison, G., Boreham, P.: 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 (2019). DOI 10.1109/CSE/EUC.2019.00061. URL https://ieeexplore.ieee.org/document/8919563/
36. Smith, C.: Remote Control Modes – Local vs Remote (2013). URL https://blog.se.com/machine-and-process-management/2013/12/17/remote-control-modes-local-vs-remote/tle
37. Steiner, W., Heise, P., Schneele, S.: 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 (2014). DOI 10.1109/DASC.2014.6979419. URL https://ieeexplore.ieee.org/document/6979419
38. Vahidinasab, V., Tabarzadi, M., Arasteh, H., Alizadeh, M.I., Mohammad Beigi, M., Sheikhzadeh, H.R., Mehran, K., Sepasian, M.S.: Overview of Electric Energy Distribution Networks Expansion Planning. IEEE Access (2020). DOI 10.1109/ACCESS.2020.2973455
39. Wang, H., Zhang, X., Sun, Y., Li, J., Li, Y.: 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 (2019). DOI 10.1109/ITAIC.2019.8785704
40. Wang, T., Wang, K., Li, J., Yu, H., Shuai, W., Bian, J., Zhao, X.: Fast recognition of human climbing fences in transformer substations. In: 2017 Ninth International Conference on Advanced Computational Intelligence (ICACI), pp. 195–200. IEEE (2017). DOI 10.1109/ICACI.2017.7974508. URL http://ieeexplore.ieee.org/document/7974508/
41. Wojke, N., Bewley, A., Paulus, D.: Simple online and realtime tracking with a deep association metric. In: ICIP, pp. 3645–3649. IEEE (2017)
42. Zhang, S., Staudt, E., Faltemier, T., Roy-Chowdhury, A.K.: A Camera Network Tracking (CamNeT) Dataset and Performance Baseline. In: 2015 IEEE Winter Conference on Applications of Computer Vision, pp. 365–372. IEEE (2015). DOI 10.1109/WACV.2015.55. URL http://ieeexplore.ieee.org/document/7045909/
43. Zhang, Y., Wang, C., Wang, X., Zeng, W., Liu, W.: Fairmot: On the fairness of detection and re-identification in multiple object tracking. arXiv preprint arXiv:2004.01888 (2020)
44. Zivkovic, Z.: Improved adaptive Gaussian mixture model for background subtraction. In: International Conference on Pattern Recognition, vol. 2, pp. 28–31 (2004)