Original research article

VIEXPAND – Industrial Video for Expanded Vision in Remote Operations

C. Ribeiro[a,b,c] (iD) 0000-0001-7484-6779, J. Gil[a] (iD) 0009-0008-2366-5669,

X. Bento[a] (iD) 0009-0000-0610-5346, M. Figueiredo[b,c,*] (iD) 0000-0002-0780-3725,

J. Rosa[b,c] (iD) 0009-0000-5543-7814, R. António[b,c] (iD) 0009-0008-6503-2739,

L. Ferreira[b,c] (iD) 0000-0003-0648-6067, P. Assunção[b,c] (iD) 0000-0001-9539-8311

[a] Twevo, Lda., Coimbra, Portugal;
[b] Polythecnic of Leiria, Leiria, Portugal;
[c] Instituto de Telecomunicações, Portugal

ABSTRACT

Manufacturers are always looking for new ways to minimize breakdowns and failures to improve productivity and efficiency of the production process. VIEXPAND is a multiview AI-enhanced video processing and transmission system, operating in real-time, for industrial supervision, inspection, surveillance and security applications, on the road to Industry 4.0. It focuses on professional applications where remote monitoring of large industrial areas requires wide fields of view and multiple views, demanding for efficient data transmission, in real-time. This paper covers a use case in a glass bottle production industry to reduce raw material use and energy waste, increase production efficiency and improve safety of operations.

1. Introduction

Real-time smart supervision involves using advanced technologies, data analytics, and automation to monitor, control, and optimize industrial processes. This is aligned with the fundamental principles of Lean Manufacturing [1] - reducing waste, increasing efficiency, and continuously improving processes.

First, it allows for immediate detection of inefficiencies and waste, enabling quick response and adjustments to maintain smooth flow and reduce delays. Second, it reduces the need for manual interventions, minimizing human error, and ensuring consistent and predictable process execution. Finally, it can predict equipment failures before they occur, allowing for proactive maintenance, reducing downtime and ensuring that machinery operates at peak efficiency.

This paper describes VIEXPAND, a real-time smart supervision solution and its application in the bottle container manufacturing industry. It was developed by two co-promoters, TWEVO, Instituto de Telecomunicações, and an industrial partner – Vidrala S.A. The main goal was to create new video acquisition and processing processes to enable automation mechanisms to be implemented in different industrial sectors, in real-time, for remote supervision, inspection, surveillance, and security applications. The project particularly focuses on professional applications where monitoring large industrial areas requires wide fields of view (up to 360°), and thus guarantee Quality of Operations (QoOs) in the distance. Such a smart industrial tool allows expanding the view of the operator, distant from the factory floor, increasing labour efficiency and yielding better productivity [2].

2. Theoretical foundations

2.1 VIEXPAND architecture

In the glass container industry, Individual Section Machine (ISM) real-time smart supervision offers opportunities to predict and identify malfunctions, reduce waste (material and energy), automatically collect and analyse production data and remove operators from an extremely harsh environmental, where human resources face difficult working conditions. The relevance of this use case is evidenced by the existence of two other commercial solutions from AVACON and CERRION. Both offer real-time supervision of ISM operation to detect malfunctions and generate alarms. However, VIEXPAND's superior video coding pipeline and heterogeneous hardware platform allows it to process up to four Full High Definition (FHD) video streams per edge device. This means it is possible to extract and intelligently combine information from multiple cameras in real-time, support event/product traceability in long production lines, and offer multi-site or multi-factory supervision. Table 1 compares VIEXPAND solution with its main competitors in this vertical. Note, that being three commercial solutions, not every technical detail and Key Performance Indicators (KPIs) are (can be) disclosed.

The VIEXPAND system is built around two major blocks, the Capture and Transmission Centre (CTC) and the Monitoring and Control Centre (MCC), connected through the cloud. The former is placed on the production site and the latter on the inspector/user site. Several CTC devices can be deployed in parallel in the same or different sites, each of which may include several video capturing devices (at least 4). Figure 1 depicts this arrangement, where, additionally, local and remote viewing displays can be deployed, as per request. The system is designed to be highly flexible: a) enabling real-time control of the encoded data rate, to dynamically adapt to the available bandwidth; b) using multi-standard video coding/decoding in transmitters and receivers; c) supporting standard wired or wireless network technologies; d) enabling dynamic video compression in Regions of Interest (ROIs); and e) supporting the integration of Artificial Intelligence (AI) technologies to detect objects and extract useful production related information.

Field-Programmable Gate Arrays System-on-Chip (FPGA-SoC) devices are used to implement the CTC, including the AI engine used for object clas-

Table 1. VIEXPAND compared to commercial solutions from AVACON and CERRION

Criteria	VIEXPAND AI by TWEVO	Solution by CERRION	Sentinel by AVACON
Performance monitoring	✓	✗	✗
Production statistics data collection & reporting	✓	✗	✗
Detect anomalies and issue warnings	✓	✓	✓
Multi- (camera, display, site, factory)	✓	≈	✗
Seamless integration on ISMs	✓	✓	✗
Edge processing capability in Frames per Second (FPS)	> 100	> 25	Unknown
Latency	< 600 ms	< 2 s	Unknown
Market history/phase in glass container industry	Short	Short	Mature
Modular for easy and less costly post-setup support	✓	✓	✗
Product lifetime support service and warranty	✓	Unknown	✗

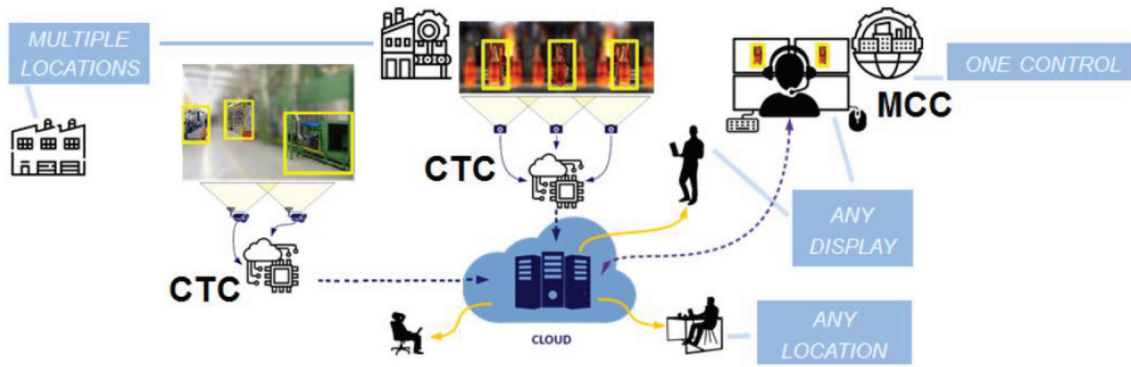Note. ✓ yes | ✗ No | ≈ to a certain extent

Figure 1. VIEXPAND overall architecture

sification, detection and tracking in the **VIEXPAND** AI solution. This heterogeneous hardware technology allows us to distribute processing tasks over different modules available on-chip (multi-processor units, Graphics Processing Unit (GPU), video coder/decoder, hard-cores and soft-cores), thus enabling real-time operation with very low latency. The CTC is responsible for: a) capturing video from multiple high-speed raw cameras or network cameras; b) conditioning and multiplexing the video streams; c) performing multiple processing operations on each video stream; d) performing ROI-aware dynamic video compression; and e) transmitting the encoded video streams to the cloud via a standard network interface. To enable further adaptability to network conditions the High Efficiency Video Coding (HEVC) [3] video encoder was enhanced to support Quantisation Parameters (QPs) and ROIs as input parameters, while the decoder is used without any modification to ensure that the stream produced by the enhanced encoder is HEVC compliant. Figure 2 presents the

CTC functional blocks and how they interact. The solid contour lines correspond to the major VIEXPAND R&D contribution (blue blocks), and dashed contours (grey blocks) are Custom-Of-The-Shelf (COTS) modules. Developed modules include a decoder to interface the camera's encoder, video multiplexing and demultiplexing, several pre-processing modules for the video streams, a module responsible to superimpose ROIs to the video streams and several control units (for transceiver throughput, video encoder and CTC). Concerning the physical interface to the camera module, VIEXPAND supports the Mobile Industry Processor Interface (MIPI) Camera Serial Interface (CSI), a high-speed point-to-point interface.

At the MCC, the video signals are decoded and displayed on a visual platform to an operator. This platform provides a setup for the operator to define a wealth of operational parameters, including the manual or automatic definition of a) the K views among the N possible video streams in real-time; b)
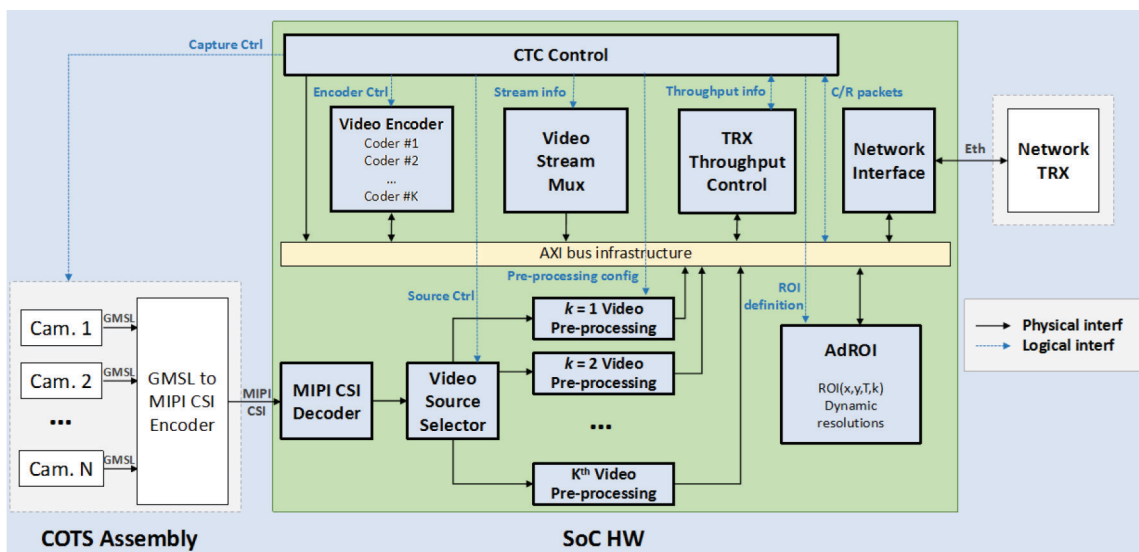

Figure 2. CTC block diagram

the video streams' quality and size; or c) the ROIs. Figure 3 presents the MCC block diagram, following the same colour scheme used above in the CTC block diagram. Developed modules include the user interface, the module responsible for the manual ROI definition, throughput monitor and control, packet generation, video demultiplexing and post-processing functions. Here, the video HEVC decoder is used without modification and thus, it is marked as a COTS module (dashed contoured grey block). This is a softcore Intellectual Property (IP) module, freely available.

2.2 Rate control algorithm

Video encoding involves the compression of video data to decrease its size without significantly compromising its visual quality [4]. Within this framework, Rate Control (RC) emerges as a pivotal element of video encoding, with the goal of aligning the compressed video's Bit Rate (BR) with a specified target BR, while upholding a satisfactory level of visual quality. Traditionally, RC algorithms are designed for general purpose applications, such as video streaming, broadcasting and storage [5] but its performance can be improved if tailored to the specific application scenario. For example, video applications with low-mobility recorded in environments with significant noise, like industrial settings, entail distinct RC requirements when contrasted with high-mobility video scenarios, e.g. sports footage [6].

In VIEXPAND, an algorithm was developed for low-mobility video applications captured in high-noise industrial settings using a fixed camera [7]. This algorithm implements two stages: first, a one-time start-up procedure is performed to find the initial RC algorithm parameters; then, the video is encoded using a QP such that the stream BR matches the target

BR. Rather than selecting the frame's QP with an analytical model, our operational approach follows the RC trends [8] by using QP-Frame Size (FS) matching function which considers the previous frame's encoded sizes. Experimental results have shown that the proposed algorithm can reach the target BR in a short period of time. In steady state, the objective visual quality Peak Signal-to-Noise Ratio (PSNR) achieved by both the reference and the proposed RC algorithm is similar [7]. Therefore, the proposed algorithm is a promising solution for industrial/surveillance applications.

2.3 ROI-aware coding methodology

In HEVC it is possible to define different QP within an image, because this parameter is defined for each Coding Unit (CU). In VIEXPAND, the source code of the HEVC reference encoder was changed to accept a lower QP in the ROIs. With the coordinates of ROIs, the edge encoder can compress (intra-coding compression) the image giving more importance to the regions relevant to the task, saving bandwidth without compromising performance [9]. ROIs can be manually defined by the human operator in the MCC and transmitted back to the CTC, or automatically determined by the AI-engine, which can be implemented in the MCC or CTC (see Section 3.2).

3. Materials and methods

3.1 Application scenarios

Two different application scenarios were identified as initial use cases. The first scenario is the glass bottle production line supervision, where ISMs pro-
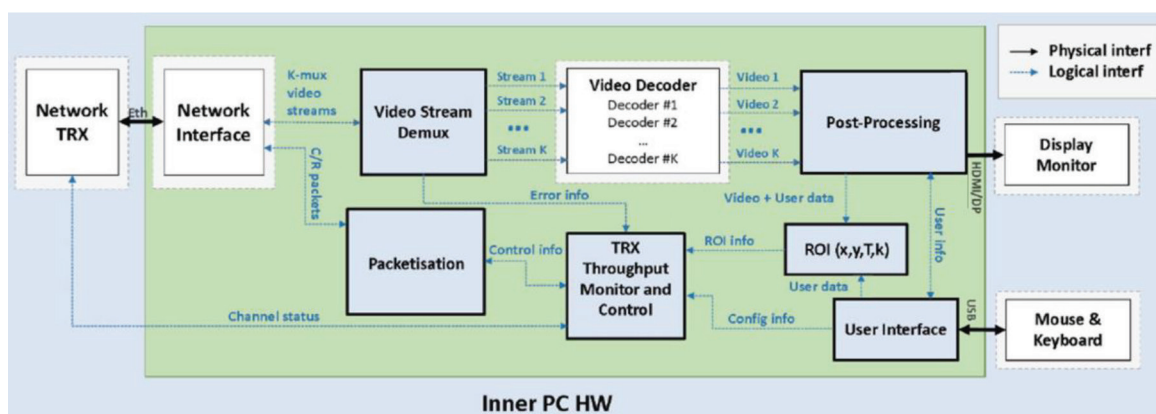

Figure 3. MCC block diagram

duce bottles out of the melted glass gob and place them onto a production line conveyor belt. This is an extremely hot and noisy scenario, with a high degree of dust and oil, where the production rate can reach 400 bottles/min at a displacement speed of 40 m/min. In this harsh environment, human workforce is hard to find and quite expensive. Figure 4a) shows the production line with camera placement and viewing angles. VIEXPAND is to function as an additional supervision control system, complementary to the current human inspection of the bottle line-up, to timely and duly detect bottle jams, fallen, misplaced or misaligned bottles and allow a faster reaction to process deviations. The second scenario corresponds to a cool, silent and low mobility environment. This factory area is where bottles are thoroughly inspected and packaged for external delivery. VIEXPAND is to function as an additional safety distance supervision control system, keeping personnel sufficiently away from forklift trucks and autonomous railed robots (Unmanned Ground Vehicle (UGV)). There is an area surrounding the ~150 m long rail, used by either UGVs, forklift trucks and walking staff. Figure 4b) shows an example of such operating area, with the possible positioning of cameras with their optical Field-of-View (FOV). Most accidents happen between the forklift trucks and UGVs, and with the trucks hitting walking personnel, mainly at the back and side of the forklift's drivers.
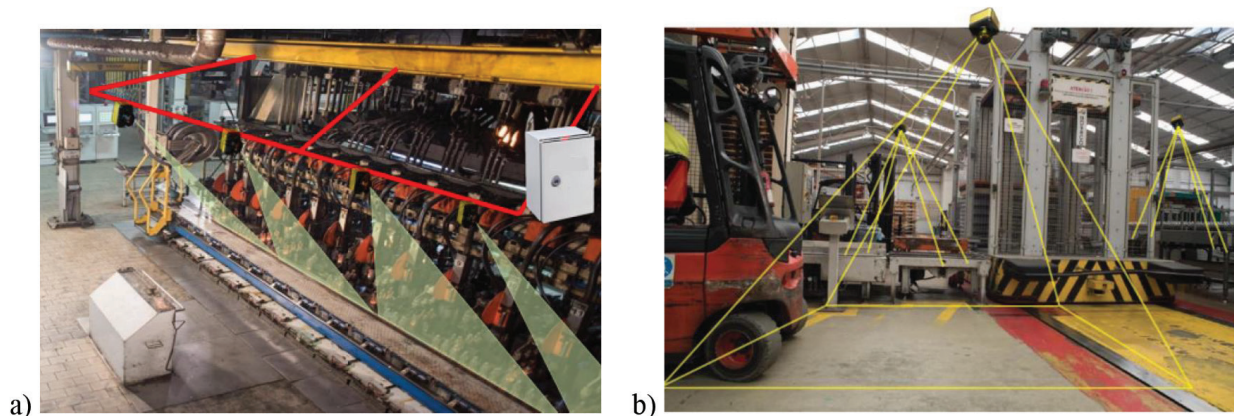
3.2 AI integration in VIEXPAND

In low-mobility video scenarios, ultra-low latency is not a crucial requirement and for that reason, AI does not need to be implemented on the edge. Edge devices need only to capture video, encode and multiplex the streams to send over the network to the MCC, where AI techniques are used to automatically extract relevant information from the received and decoded video streams, as well as the definition of the ROIs to send to the encoder (Figure 5a). During the start-up period, the CTC has no ROI information to perform ROI-aware coding and thus, full quality images must be transmitted to the MCC to enable accurate initial ROI identification. However, after this initial configuration, the overall transmission bandwidth can be reduced by reallocating resources (bits per pixel) to ROIs. This can substantially decrease the overall steady-state transmission bandwidth due to the possibility of staggered timing in the initial set-up of multiple sites.

Traditional video compression standards have a significant impact on the performance of AI algorithms [10], because machines do not have the same perceptual biases as humans [11]. For this reason, VIEXPAND implements ROI-aware coding, giving a higher priority to ROIs in the compression process, meaning that more bits are allocated to these regions and are therefore more likely to be accurately represented in the compressed video. In this remote AI-assisted scenario, ROIs are no longer defined by the human controller in the MCC - the AI engine automatically identifies those regions and reports them back to the encoder (the round-trip delay of ROI information can be considered negligible in low-mobility scenarios). If the AI engine is implemented in the CTC, this start-up procedure is not required because ROIs are available at the encoder.

In high-mobility scenarios, such as the bottle production line, the AI engine should be placed as close as possible to the video source, i.e., at the CTC (see Figure 5b)). As in the previous scenario, the AI engine implements a Neural Network (NN) which is responsible for detecting the objects of inter-

Figure 4. VIEXPAD application scenarios with the indication of the camera placement and their FOV: a) glass bottle production line; and b) UGV and forklift truck operating area
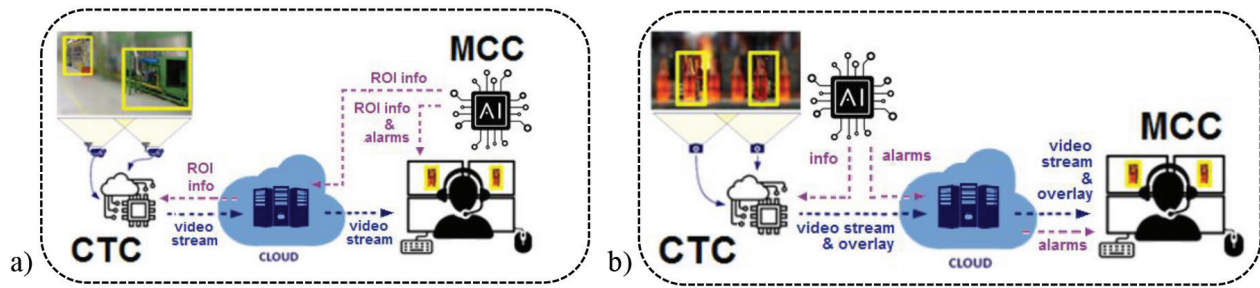
Figure 5. AI integration: a) at the MCC for low-mobility video; b) at the CTC for ultra-low latency

est (instance or semantic segmentation NN can also be used, depending on the application) and passing its position information to the inspector entity. The inspector identifies anomalies and generates alarms. This bundle of information is then provided to the AdROI entity (shown in Figure 2)) that performs 2 main tasks: 1) generates the operational overlay that is superimposed to each of the transmitted video streams; and 2) generates the Bounding Boxes (BBs) around the objects in the multiple processed video streams that will used by the encoder to compress the images, giving better quality to the relevant regions.

3.3 Edge AI implementation solutions

The FPGA MPSoC used to implement the CTC (Xilinx Zynq UltraScale+ [12]), supports a soft-core Deep-Learning Processing System (DPU), capable of implementing several YOLOX models in parallel [13]. This enables real-time object detection in multiple video streams, in a single CTC. You Only Look Once (YOLO) family of object detectors are state-of-the-art object detectors for real-time applications [14]. It comes with a wide range of size options with implications in latency, hardware footprint and energy consumption.

In VIEXPAND AI, several YOLOX networks are used in parallel, one for each video stream. The raw video data is pre-processed before entering the NN to adapt the image format, size and pixel gain to its requirements. In the resizing operation, the input size aspect ratio is maintained similar to the original image size (to minimize input distortion) but reduced to 640x352 pixels. This smaller size has proven to have the better trade-off between complexity and accuracy, with a mean Average Performance (mAP) of 55.4%, measured at an Intersection over Union (IoU)=0.5:0.95, for a maximum number of detections equal to 100 and within 3000 training epochs [15]. Complexity was evaluated both in terms on network parameters (5.03 Million) and operations (8.13

Giga Operation Per Second (GOPS)), to process 4 video streams with 30 FPS each.

For NN training, a specialized dataset was generated utilizing an extensive collection of images captured along the bottle production line and in different sites of the factory. This dataset encompasses a diverse array of bottle types, with various shapes, colours and orientations. Labelling was performed with an AI-based online tool, after which all labels have been manually verified, deleting any false positives (for instance, the shiny steel background moulds with the shape of a bottle) and adding labels on missed bottles. This methodology allowed us to label almost 300 images, generating more than 8000 labels. While the dataset might appear relatively compact, it proved to be ample enough for the intended application where the scenario is highly controlled. Nevertheless, it is important to note that this custom dataset must be updated with new images every time VIEXPAND AI is used in a different scenario.

Experimental results have shown that these MPSoCs can be used to process multiple video streams, with reasonable detection performance, while providing enough computational resources for other tasks [15]. Different network configurations and training approaches were tested to select the best solution, given the compromise between performance and computation requirements. Following those results, it was decided to use the ZCU104 evaluation board to implement the CTC in high-mobility scenarios (where ultra-low latency is required) and the low cost KV260 Vision AI Starter Kit in low-mobility scenarios. The first uses a mid-range device that supports dual B4096 DPU cores [16], delivering 2.46 Tera Operations Per Second (TOPS), 8-bit integers (INT8), peak performance for NN inference acceleration. The second uses a custom device that supports only one B4096 DPU core and delivers 1.23 TOPS INT8 peak performance.

4. Results

## 4.1 Installation and operational requirements

To operate in these different industrial scenarios, specific requirements have been established from the start. At the end of the project, all these requirements have been met and experimentally assessed. These are:

1. Guarantee optimised video quality by implementing lightweight dynamic and adaptive algorithms for control and parametrisation of functions, as well as for real-time analysis of the transmission, reception, Bit Error Rate (BER) and throughput. In VIEXPAND, the developed RC algorithm has shown to achieve a similar visual quality (PSNR), compared to the reference RC algorithm in HEVC, but with a faster response and lower implementation complexity, as reported in [7].

2. Implement dynamic ROIs, determined manually by the system's operator, automatically by AI methods (where the area of the detected objects become the ROIs), or both (the operator may define the ROI where AI algorithms are applied to detect objects). Regardless the method, efficient coding should be adapted to both human and machine vision demands. In VIEXPAND, the efficient object/ROI detection and coding algorithms allowed us to reduce the bitstream up to 40%, in comparison with the HEVC standard, with a similar accuracy in the object detection task [10]. Also, we were able to reduce the bitstream up to 50%, without a perceptible visual quality degradation, as described in section 4.2.

3. Build the system in low size, weight and

power (SWaP) FPGA-SoCs, supporting real-time software and firmware, with sub-second end-to-end latency and frame rates between 30 and 60 FPS. In VIEXPAND, we implemented all video pipeline and AI engine in a low SWaP device, as described in section 2.1 and reported in [15], with the ability to process 25 FPS per stream (4 video streams) in ~550 ms, as described in section 4.3.

TWEVO's VIEXPAND AI solution has already reached Technology Readiness Level 6 (TRL6) in the glass bottle production sector. A pilot has been installed in the Vidrala Gallo Vidro factory, in Marinha Grande, Portugal, in both scenarios: monitor ISM #5.4 machine (scenario 1); and monitor people alongside running forklift trucks and UGVs (scenario 2). In both, we resorted to CTCs with AI engines implemented on edge devices, following the architecture illustrated in Figure 5b) although scenario 2 could have been implemented with AI in the MCC. In scenario 1, the CTC was mounted inside an industrial control panel enclosure, together with steel enclosure boxes for each camera, and flexible PVC-lined metal duct hoses to protect the camera cables (keeping in mind that this area is extremely hot). Further, a ventilation air inlet valve has been added to control the cooling airflow into the boxes and hoses. Figure 6a) shows the top view of the main CTC enclosure, with the Xilinx ZCU104 Evaluation Kit, the AES-FMC-MULTICAM4-G FMC module, power supply and auxiliary solid-state drive. Figure 6b) presents the detail of the camera enclosures, with each lens, AR0231AT Image Sensor Board and MAX96705 Serialiser Kit.

Installation is depicted in Figure 7a), showing the approximate FOVs of the 4 cameras, towards the conveyor belt. On the MCC side (Figure 7b)), the system is fundamentally hosted on a mini-PC with a user-
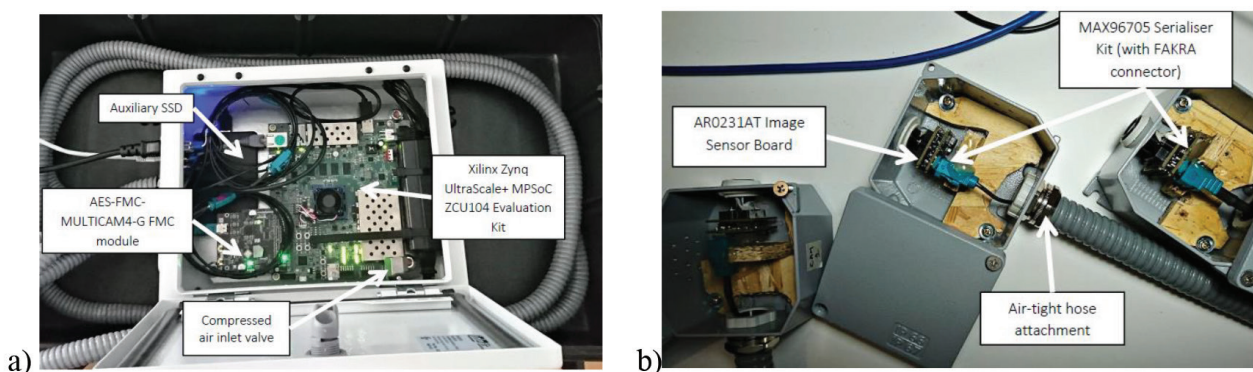
Figure 6. Image of: a) CTC block, with the enclosure open; b) cameras, fit inside the protective metal enclosures
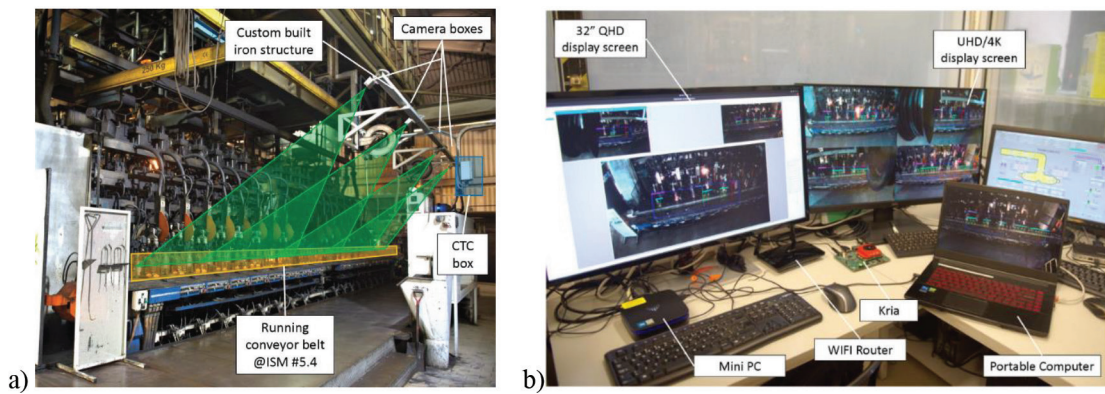
Figure 7. Image of: a) the ISM machine, in Scenario 1, with indication of the cameras' approximate FOVs; b) the Control Room, where the MCC is placed, and several displays are being used for visualisation

friendly Graphical User Interface (GUI) and a Kria board to enable low latency decoding and visualization (see section 4.3). We can also see a PC acting as an additional visualization unit, a WIFI router and several display screens. After installation in such hard temperature and moist environment, the system has shown excellent operation stability during long term tests.

In scenario 2, three cameras were installed and directed towards the running UGV, as shown in Figure 8. The CTC was built around a Xilinx Kria KV260

Kit with additional network switching, power, cooling and connections. All 3 cameras are different network models, two from maker Dahua and one from Uniview, demonstrating the flexibility and interoperability of VIEXPAND system in adopting different camera setups. The MCC was implemented in a portable PC for control and visualisation. By visually merging the images from the 3 cameras (shown in Figure 9), we exemplify how effectively the visual perception is expanded, covering a large area within the warehouse.
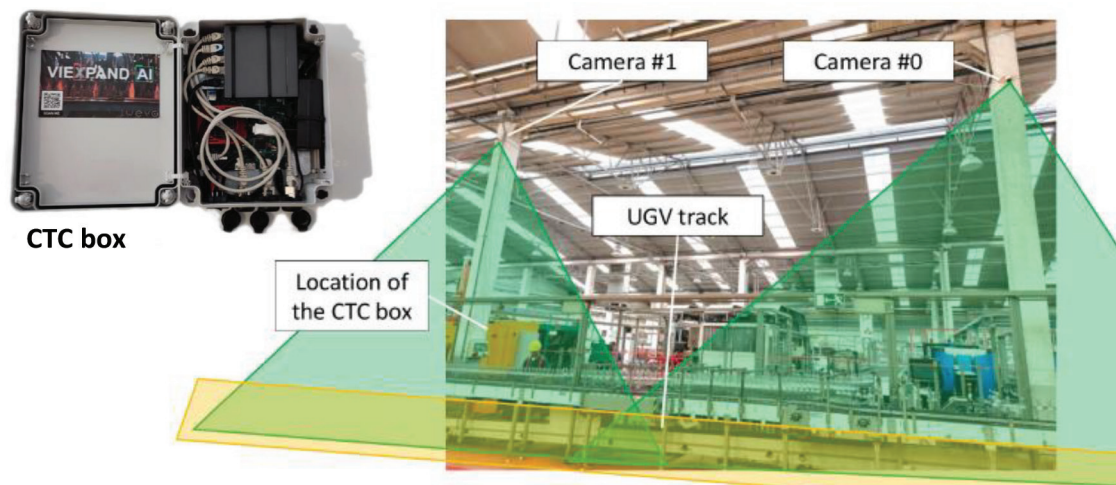
Figure 8. Image showing cameras #1 and #0 installed, with the indication of the approximate FOVs and UGV track

Figure 9. Detection of people within the ROI defined by the user, covering the combined 3 camera FOVs

4.2 Visualization requirements

An enormous gain this solution comes from the ability of using AI models on smaller size images, to detect the objects of interest, apply the dynamic ROI, and translate the ROI information back to the original full scale 1920x1080 pixels FHD image. This FHD video, with the operational overlay on top, will then be encoded and transmitted in real-time to the visualization devices. This means that, unlike other AI applications in the market that reduce the image size and sacrifice final human visual quality, VIEX-PAND is benefiting from its real-time capability of best visualisation for the human, while keeping the dynamic adaptive ROIs. To show the implications of this, Figure 10 shows a detail of the same image, with bottles on the running conveyor belt, in full scale 1920x1080 pixels FHD, on a FHD screen and in real-time (cropped to 50% in height for space saving purposes). Alternatively, Figure 11 shows the same detail of bottles in reduced scale 512×288 pixels put then up-sampled back onto a 1920×1080 pixels FHD

screen for human visualisation (here also cropped to 50% in height). This is the case of most AI real-time implementations but resulting in enormous visual degradation.

We have gone forth with two test situations (scenario 1) to evaluate the relation between throughput, transmitted data, and visual perception, after the user defines a ROI and its quality (QP offset). First, we compare overall image quality, final whole visual perception and user satisfaction with constant Bit Rate (BR) - ROI control priority follows a specific set bitrate, resulting in consequent base quality given by an unknown final QP. Figure 12 and 13 shows the results for BR=300 kbps and a QP offset set to 0 or 11, respectively. We can see no quality difference in Figure 12, but in Figure 13, the image quality outside of the ROI (detection area defined by the blue box) is degraded and improved inside the ROI. So, the resulting quality in the ROI of Figure 13 is better than in the ROI of Figure 12, while keeping the same BR.

We also compared overall image quality and final whole visual perception with variable BR - ROI


Figure 10. Detail of bottles in full scale 1920×1080 pixels FHD, on a FHD screen (cropped 50% in height)

Figure 11. Detail of bottles in small scale 512×288 pixels on a FHD screen (cropped 50% in height)

control priority follows a specific set quality, resulting in a consequent final bitrate. In Figure 15, QP offset is set to 10, while the bit rate is reduced when the background quality is degraded, i.e., the resulting bit rate in Figure 15 (BR=1.010 Mbps) being much lower than in Figure 14 (BR=2.169 Mbps) for the same ROI quality. Note that the bottles and buttons outside the ROI have much lesser definition in Figure

15, while the quality inside the ROI is very similar. This means that the sacrifice in overall quality (background degradation), in favour of better ROI visual quality and same bit rate, does not significantly result in user dissatisfaction or reduction in QoOs (Figures 12 and 13). Also, the operator can hardly discern any differences between Figures 14 and 15 yet resulting in a significant bitrate reduction of approximately 50%.
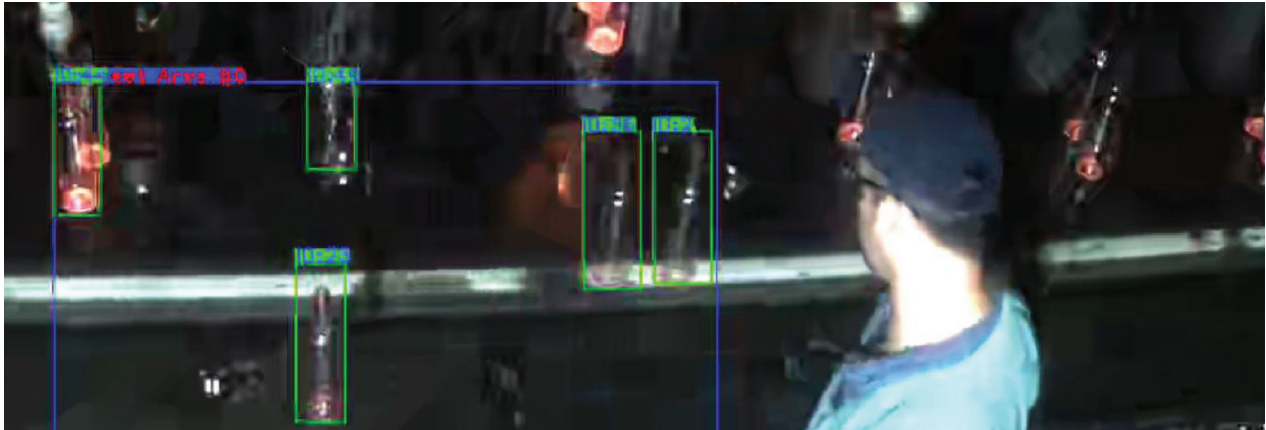
Figure 12. With no QP offset and BR=300 kbps, there is no quality difference between ROI and remaining image

Figure 13. With a QP offset of 11 and BR=300 kbps, there is a quality difference between ROI and remaining image
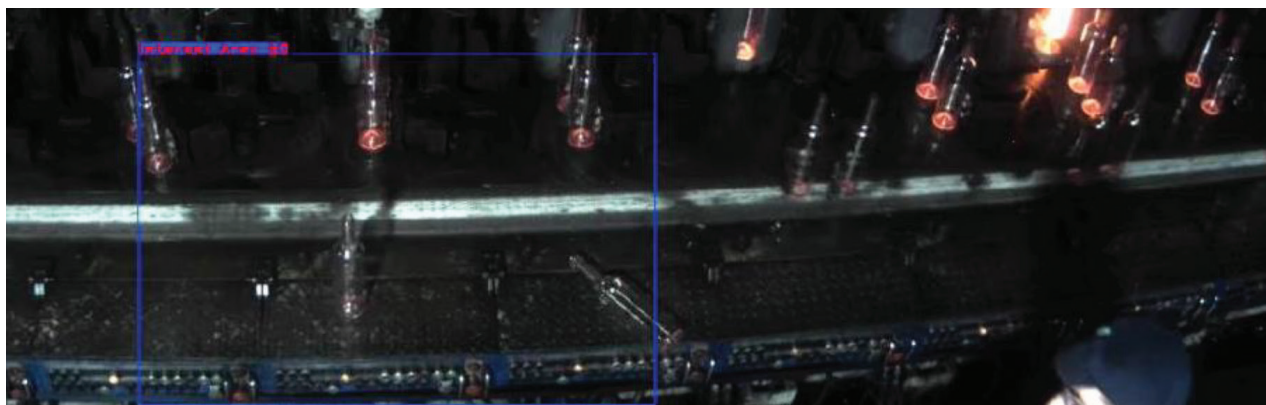
Figure 14. Quality difference between ROI and remaining image, with no QP offset, resulting in a BR=2.169 Mbps
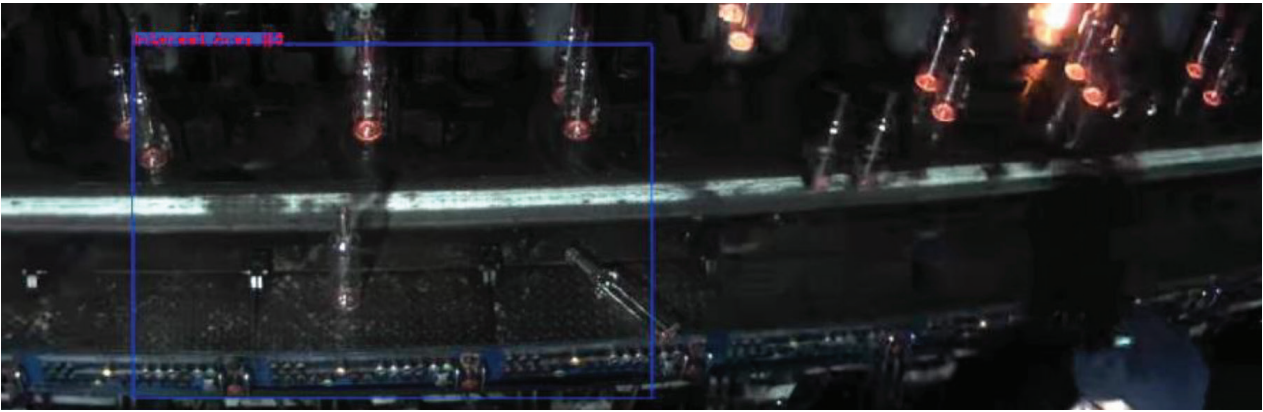
Figure 15. Quality difference between ROI and remaining image, with QP offset=10, resulting in a BR=1.01 Mbps

4.3 Streaming video latency

Streaming tests were executed with 1920×1080 pixels 30 FPS network cameras, for both Real Time Streaming Protocol (RTSP) and User Datagram Protocol (UDP) streaming. These tests involved the successful coding and output of several streaming video files, with and without ROIs, either at 128 kbps or 5 Mbps data rates, using the setup depicted in Figure 16. One of the video cameras recorded the current time in a monitor displaying a chronometer, and all 4 camera streams were transmitted to the MCC and visualized in a second monitor. A camera was then used to take multiple synchronized photographs of both monitors, which allowed the extraction of the system's latency mean and standard deviation values. Standard deviation (stdev) was observed to be around 33 ms, which corresponds to the frame capture delay. Wireless links have also been tested but introduce a higher latency that significantly depends on the quality of the transceivers, and for that reason, we decided to use Ethernet connections.

We noticed that using a standard PC to decode streams in the MCC side results in a significantly high latency (>1 s). This increased latency is highly dependent on the actual machine used (joint operation of CPU, GPU, RAM and operating system, also dedicated to other miscellaneous operations). The tested machines make use of low-end video cards/GPU, most likely being the most important culprit for the large latency obtained. To circumvent these issues, we implemented a visualizer in a high performance, dedicated video decoder board (using the same Xilinx KV260 Kria board we used for the CTC in scenario 2) directly interfacing with the output display.

Table 2 summarises the average and standard deviation results for all the Kria RTSP measurements (Kria UDP measurements showed similar results). We see that, overall, total latency does not considerably vary with data rate, number of ROIs or active streams. Latency variation has tendencies that are justifiable by the following reasoning:

- Lower data rate implies larger latency, especially with lower number of ROIs, lower number of active streams and for lower bit rates. These are expected results, since lower data rates and lower number of ROIs (larger area of the im-
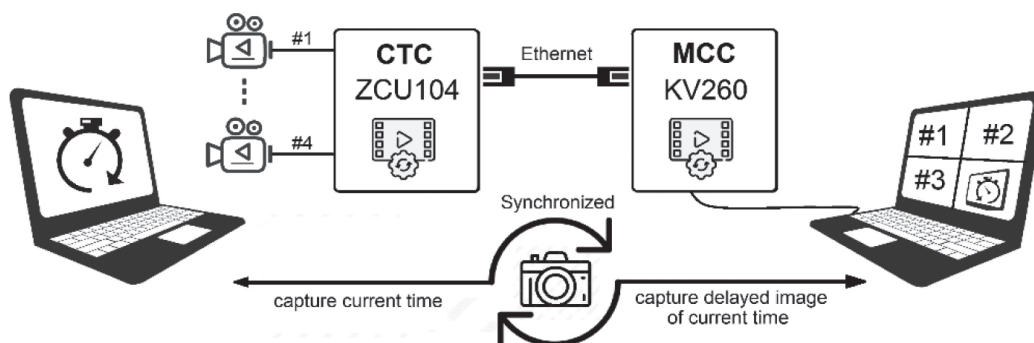

Figure 16. Setup for video latency streaming tests

age to be coded with lower quality/higher QP) imply larger coding effort.

- The number of streams has minimal impact on the latency showing that the system is very efficient in pipelining the coding/decoding operations.

With the AI engine running on the CTC, similar tests were conducted with RTSP. We used Kria boards on both CTC and MCC side, and standard IP cameras, to evaluate the worst-case scenario (low-cost Edge devices with standard industrial cameras). These cameras impose an additional processing time, compared with the camera modules used before. The network IP camera captures and encodes the video stream for transmission, then transmits the video stream to the network; it is then decoded at CTC Kria, processed and coded back again for transmission to the MCC Kria, which receives the stream, decodes it and outputs the stream to the display. Tested video transmission rates are again 128 kbps and 5 Mbps, with a maximum of 50 detectable BBs and 4 active streams. Results have shown average latencies of only 557 ms (stdev=36 ms) and 572 ms (stdev=28 ms), respectively, for 128 kbps and 5 Mbps, keeping the frame rate at 25 FPS. This is only a small latency increase (~200 ms) in respect to not having any AI engine working, showing the efficiency of implemented AI algorithms.

5. Conclusions and future work

The project final prototype is now able to handle 3 types of ROI definitions (as required, and explained in section 4.1), providing flexible and dynamic ROI definition for the operator. It is also able to deal with bottle, persons and faces detection, particularly adapted for the project target scenarios: inspection of bottle production line; and supervision of staff safety around UGV and forklift trucks (or other restricted

areas). Concerning operation, factory staff has understood how the system operates, from the selection of ROIs on the MCC control display to the extracted measurements and issued warnings. Concerning the final subjective latency and video quality results, we have verified that delays are almost imperceptible using the Kria display and registered the very positive remarks on the perception of image quality that results from a combination of resolution, exposure and frame rate.

By the end of the project some shortcomings and difficulties were identified. First, the human manual intervention on the training process (video pre-processing, labelling, training and optimization) is significant. This can introduce scalability challenges when it comes to deploy this technology in different production lines or industrial scenarios. Second, the system relies only on supervised learning AI techniques, which cannot efficiently handle unexpected (unseen) events/objects or operational conditions. This can result in a degradation of the confidence levels used to take decisions and issue warnings. Finally, as common to all AI-based technologies, it is not always obvious how the system learns and why it makes certain decisions. For that reason, it is not easy to identify biases, errors, limitations in the learning process, and areas for improvement and optimisation.

To elevate this project to a next level of accuracy, reliability and dependability, and conquer both clients' and workers' confidence, VIEXPAND AI have had further developments. We adopted a cloud service-based model (on-going developments) to reduce manual human intervention and guarantee continuous improvement and adaptability. This service should remotely monitor the system performance, automatically retrain algorithms (with active learning and self-supervised methodologies) and promote software/hardware updates to the edge device. Enabling remote and automatic lifetime functionality improvements, also enables VIEXPAND to deliver ongoing value and stay competitive in a rapidly evolv-

Table 2. Overall stream latency results, with the Kria and RTSP

RSTP	128 kbps, 0 ROI				128 kbps, 5 ROI/stream		
Number of streams	1	2	4	% of variation	1	4	% of variation
Average latency (ms)	436	397	381	-13%	369	391	6%
Standard deviation (ms)	17	35	39		31	29	
	5 Mbps, 0 ROI				5 Mbps, 5 ROI/stream		
Number of streams	1	2	4	% of variation	1	4	% of variation
Average latency (ms)	375	367	401	7%	373	399	7%
Standard deviation (ms)	26	27	21		31	22	

ing market. As future developments, adding unsupervised anomaly detection algorithms could help the system handle unexpected situations and become more reliable. Finally, explainable AI technologies would allow VIEXPAND become cognitive, helping us and our clients to better understand the system's operation and identify areas of improvement, especially when operation conditions change, and performance degrades.

We believe that AI-powered industrial computer supervision technologies, like VIEXPAND, can significantly contribute to more sustainable and efficient industries. In glass container production lines, disruptive events in ISMs not only have an impact on the production line itself, resulting in machine downtime, but also upward towards the glass melting furnace, resulting in crucial energy waste. This relevant use case, described in this paper, allowed us to showcase the merits of this technology. However, this technology is flexible enough to be taken to other verticals. With minor software changes, algorithms can be adapted to detect defects and ensure products meet quality standards, identify bottlenecks and optimize workflows, detect hazards and ensure compliance with safety protocols, and perform inspections faster and more accurately than human workers. In fact, VIEXPAND AI solutions are currently being assessed and deployed to perform supervision and quality control in the automotive sector, intralogistics and food production verticals, helping to promote production efficiency, improve work conditions and reduce the carbon footprint.

Funding

References

[1] B. Rahardjo, F.-K. Wang, R.-H. Yeh, and Y.-P. Chen, "Lean Manufacturing in Industry 4.0: A Smart and Sustainable Manufacturing System," Machines, vol. 11, no. 1, p. 72, 2023, doi: 10.3390/machines11010072.

[2] Ahmad H. M. and A. Rahimi, "Deep learning methods for object detection in smart manufacturing: A survey," Journal of Manufacturing Systems, vol. 64, pp. 181-196, 2022, doi: 10.1016/j.jmsy.2022.06.011.

[3] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," IEEE Trans. on Circuits and Systems for Video Technology, vol. 22, pp. 1649-1668, 2012, doi: 10.1109/TCSVT.2012.2221191.

[4] A. A. Ramanand, I. Ahmad and V. Swaminathan, "A survey of rate control in HEVC and SHVC video encoding," 2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Hong Kong, China, 2017, pp. 145-150, doi: 10.1109/ICMEW.2017.8026268.

[5] B. Li, H. Li, L. Li and J. Zhang, " λ Domain Rate Control Algorithm for High Efficiency Video Coding," IEEE Transactions on Image Processing, vol. 23, no. 9, pp. 3841-3854, 2014, doi: 10.1109/TIP.2014.2336550.

[6] J. Zhang, S. T. W. Kwong, T. Zhao and H. H. S. Ip, "Complexity Control in the HEVC Intracoding for Industrial Video Applications," IEEE Transactions on Industrial Informatics, vol. 15, no. 3, pp. 1437-1449, March 2019, doi: 10.1109/TII.2018.2844214

[7] J. Rosa, R. Antonio, L. Ferreira, M. Figueiredo, P. Assuncao, and C. Ribeiro, "Rate Control Method for Video Encoders Operating in Industrial Environments," in 2023 Asia Symposium on Image Processing (ASIP), Tianjin, China, 2023, pp. 128-131, doi: 10.1109/ASIP58895.2023.00028.

[8] H. Zeng, J. Xu, S. He, Z. Deng, and C. Shi, "Rate Control Technology for Next Generation Video Coding: Overview and Future Perspective," Electronics, vol. 11, 2022, doi: 10.3390/electronics11234052.

[9] H. Choi and I. V. Bajic, "High Efficiency Compression for Object Detection," in 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 2018, pp. 1792-1796, doi: 10.1109/ICASSP.2018.8462653.

[10] R. Antonio, J. Rosa, L. Ferreira, M. Figueiredo, P. Assuncao, and C. Ribeiro, "Enhanced Object Detection in Highly Compressed Images using Regions of Interest," in 2023 6th Int. Conf. on Sensors, Signal and Image Processing, Nanjing China, 2024, pp. 14-19, doi: 10.1145/3653863.3653873.

[11] K. Fischer, C. Herglotz and A. Kaup, "On Intra Video Coding And In-Loop Filtering For Neural Object Detection Networks," in 2020 IEEE International Conference on Image Processing (ICIP), Abu Dhabi, United Arab Emirates, 2020, pp. 1147-1151, doi: 10.1109/ICIP40778.2020.9191023.

[12] Xilinx. "DS890 - UltraScale Architecture and Product Data Sheet: Overview", v4.4.1. [Online]. Available : https://docs.xilinx.com/v/u/en-US/ds890-ultrascale-overview [Accessed: 27-Feb-2023].

[13] Vitis AI Library User Guide. AMD Xilinx, UG1354, January 12, 2023 [Online]. Available: https://docs.xilinx.com/r/en-US/ug1354-xilinx-ai-sdk [Accessed: 27-Feb-2023].

[14] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO Series in 2021," arXiv:2107.08430, 2021.

[15] C. Ribeiro, M. Figueiredo, P. Assuncao, L. Ferreira, J. Gil, and X. Bento "Real-time industrial machine vision supervision using DPU-based edge devices," in 4th International Conference on Computer Vision and Information Technology (CVIT), Beijing, China, 2023, doi: 10.1117/12.3015817.

[16] DPUCZDX8G for Zynq UltraScale+ MPSoCs Product Guide. AMD Xilinx, PG338 (v4.1) January 23, 2023 [Online]. Available: https://docs.xilinx.com/r/en-US/pg338-dpu [Accessed: 27-Feb-2023].