Optimal Use of Cloud and Edge in Industrial Machine-Vision Applications

An Industry IoT Consortium Whitepaper

2022-11-30

Authors

Kendra Serbinski (Labforge Inc.), Yassir Rizwan (Labforge Inc.), Eugene Chang (Toshiba America Electronic Components, Inc.), Daniel Young (Toshiba America, Inc.).

CONTENTS

1 Applications of Industrial Machine Vision 3

2 Edge Computing In Industrial Machine Vision 5

2.1 Strengths of Edge Computing 7

2.2 Limitations of Edge Computing 7

3 Cloud Computing In Industrial Machine Vision 8

3.1 Strengths of Cloud Computing 8

3.2 Limitations of Cloud Computing 9

4 Deciding Where the Edge Lies in Industrial Machine Vision 9

4.1 Industrial PC (IPC) 9

4.2 Smart Camera 10

4.3 Smart Camera + Industrial PC 10

4.4 Choosing the right solution for your application 11

5 Considerations for Machine Vision Using AI 11

6 Conclusion 12

Bibliography 12

Authors & Legal Notice 14

FIGURES

Figure 1-1: A block diagram for a typical vision system operation and applicable AI techniques. Obtained from [3]. 4

Figure 1-2: A depiction of computer vision scanning a welded area. From left to right: 1) scanning of the welding part, 2) the depth image, 3) the result of the measurement. Obtained from [5]. 4

Figure 2-1: Early Automatix machine-vision system "Autovision II" at "Technology '83" trade show in Israel. Obtained from [6]. 5

Figure 2-2: National Instruments CompactRIO ruggedized controller and camera. Obtained from [7]. 5

Figure 2-3: Parabolic distortion correction to enable OCR. 6

Figure 2-4: Latest generation smart cameras have built in capability to process neural networks. This allows for unstructured and more flexible inspections. Obtained from [8] [9] [10]. 7

Figure 4-1: Camera + GPU enabled IPC via GigE Vision or USB Vision. Obtained from [22] [23] [24]. 10

Figure 4-2: Proprietary smart cameras. Obtained from [5] [7]. 10

Figure 4-3: Smart camera + compact IPC via GigE Vision. Obtained from [25]. 11

Figure 5-1: A visual representation of an intelligent machine vision model for defective product inspection. Obtained from [4]. 12

With advances in technology, cameras and AI are becoming increasingly sophisticated. It is now possible for data processing to happen directly on the sensor, at a computer nearby, at a server on premise or across the internet in a remote data center. This paper discusses the strengths and limitations of edge and cloud computing, and their applications in industrial machine vision. It aims to be a brief guide for integrating leading machine vision and edge computing practices in industrial settings and it showcases several practical examples.

This whitepaper is structured into sections as follows:

- *Applications of Industrial Machine Vision* broadly describes how industrial machine vision is being used.
- *Edge Computing in Industrial Machine Vision* gives an overview of the strengths and limitations of applying edge computing to industrial machine vision applications.
- *Cloud Computing in Industrial Machine Vision* discusses the strengths and limitations of using cloud computing in industrial machine vision.
- *Deciding Where the Edge Lies in Industrial Machine Vision* mentions several configurations of implementing edge computing in industrial machine vision.
- The *Conclusion* summarizes observations and conclusions made herein.

1 Applications of Industrial Machine Vision

Automated inspection in manufacturing has been dramatically altered by machine vision [1]. Machine vision encompasses the methodology and technology involved in extracting data from an image or series of images to produce an output used to guide a decision-based algorithm. Machine vision requires method and expertise from manufacturers as well as the integration of many technologies, software and hardware products and integrated systems. In a simplified workflow, machine vision entails imaging, followed by automated image analysis to extract the necessary information to guide a decision (Figure 1-1). [1], [2]

Figure 1-1 illustrates a typical machine-vision system operation along with areas that could use the latest AI techniques highlighted in orange. In the case of Bottlenose™, many of these tasks, such as feature extraction, classification, detection and depth processing have been moved onto the device. See side bar for a short overview of the semiconductor technology that enables these AI techniques.

Machine vision can be used for quality control and it can detect several types of defects [4]. Examples of some quality metrics machine vision can detect are listed below [3].

- Presence/absence: checks if the object is present in the checked spot.
- Orientation: determines if the object is in the correct orientation.
- Position: checks whether the object is in the right place.
- Color: checks for the correct colors in desired areas of the object.
- Recognition/content analysis: checks codes including barcodes, letters and RFID on the object.
- Geometric control: checks the object for the correct dimensions and geometric tolerances.

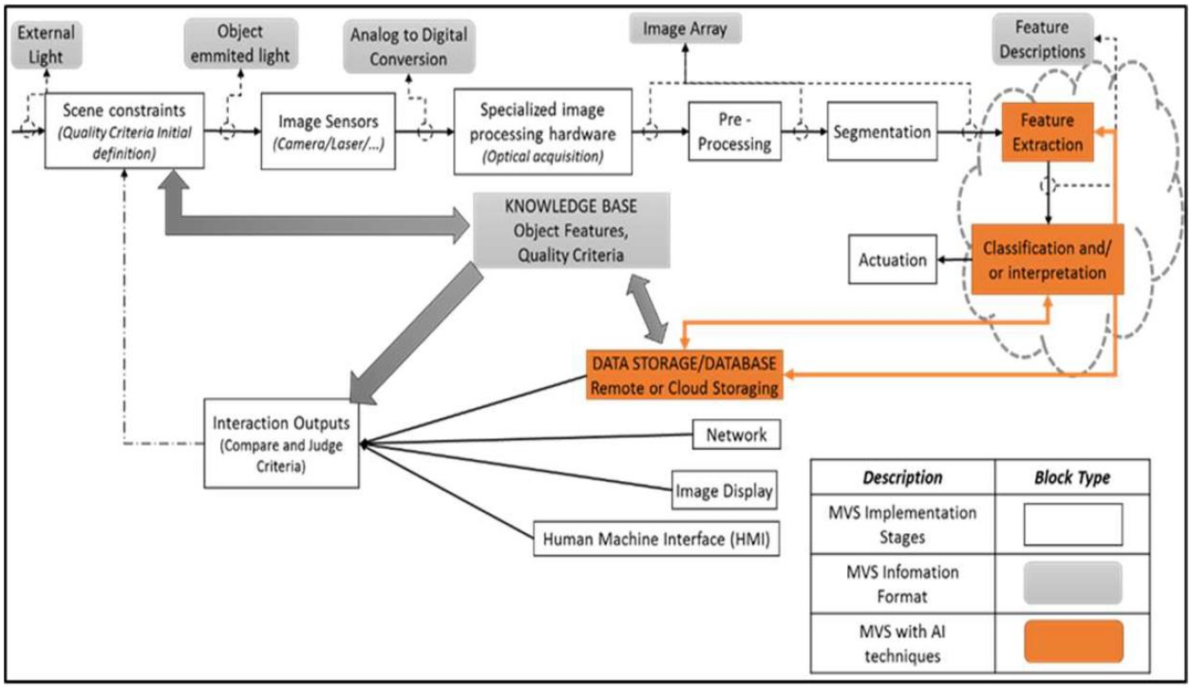

Figure 1-1: A block diagram for a typical vision system operation and applicable AI techniques. Obtained from [3].

Machine-vision algorithms typically use geometric methods with hand-defined feature descriptors. For example, a detector for the presence/absence task could be designed using contours and edges. Newer detection methods are based on DNNs that use stacked layers of convolutions and other functions that apply to the input image. Parameters of these functions can number in the millions. The DNNs are trained using large datasets and computers.

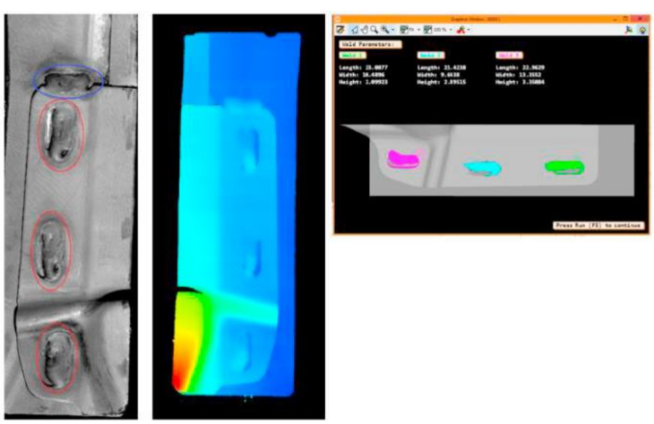

Figure 1-2: A depiction of computer vision scanning a welded area. From left to right: 1) scanning of the welding part, 2) the depth image, 3) the result of the measurement. Obtained from [5].

The state-of-the-art technology within the Labforge Bottlenose™ camera is powered by the Toshiba Visconti-5™ semiconductor device.

As Toshiba's fifth-generation image recognition processor chip, it has multiple ARM CA53 cores, multiple machine vision processing cores and an on-board Deep Neural Network (DNN) for real time AI processing power. With over 20 TFLOPS of computational power, Visconti-5 is capable of enabling vision processing with low power in a small form factor.

2 EDGE COMPUTING IN INDUSTRIAL MACHINE VISION

Industrial machine vision has traditionally used edge computing. Over the last four decades this has taken the form of a camera connected to a CPU over a communication interface like Ethernet (GigE), Firewire (IEEE 1394), Camera Link, and even USB 3.0. This CPU would reside in a purpose-built processing system, like the Autovision II by Automatix (Figure 2-1) or an Industrial PC (IPC).

Figure 2-1: Early Automatix machine-vision system "Autovision II" at "Technology '83" trade show in Israel. Obtained from [6].

Traditional computer (i.e. PC) vision systems generally comprised a lens, a camera, a frame grabber (a specialized circuit board designed to capture or "grab" individual images from the camera and send them to the computer as a single still image), a computer and machine vision software that analyzes the still images (preferably in real or near-real time). As computers and camera technology progressed, the form factors became smaller and frame grabbers were largely integrated into cameras (as shown in the GigE Vision camera in Figure 2-2), enabling the edge computer to be located closer to the camera itself and in more rugged environments.

Figure 2-2: National Instruments CompactRIO ruggedized controller and camera. Obtained from [7].

As computing power increased, so too did machine-vision algorithms and tools, but image capture, transmission and vision processing on an edge computer remained the same.

For example, in manufacturing, machine vision originally included basic algorithms, including:

- pattern matching that compared a portion of an image to a pre-defined template and calculated a correlation result (typically used to give a pass or fail judgement),
- edge detection that could be used to count items (such as the number of sheets in a roll of tissue by counting the perforated edges),
- defect detection, such as scratches on a polished surface and
- distance measurements between points of interest (such as detected edges).

Eventually, some machine-vision cameras (a precursor to today's smart cameras) could perform simple pattern matching by themselves, but their limited capabilities and relative cost limited their use to high-rate factory processes that handled a high number of products per cycle.

With the advent of optical character recognition (OCR), machine vision began to be used more widely in industry. Vision systems could now recognize alphanumeric characters, and this enabled machine vision to perform a wider variety of tasks, such as reading serial numbers and testing the functionality of closed captioning in a television manufacturing plant, as in Figure 2-3.

As with other algorithms used in machine vision, OCR has also now moved to neural network methods.

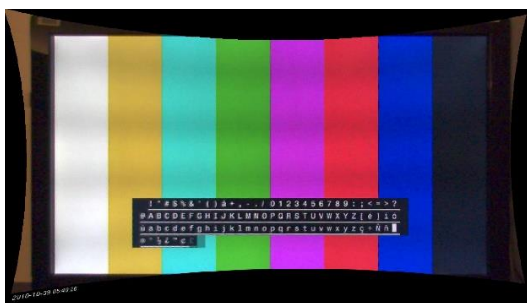

Figure 2-3: Parabolic distortion correction to enable OCR.

The last decade has seen growing interest in cameras that can do all of the machine vision processing on the device itself. Some of these cameras are shown in Figure 2-4. This is also considered edge computing. Labforge's Bottlenose cameras can do most of the machine vision tasks on the camera itself, but can also operate in a hybrid setting where parts of the front-end processing are offloaded onto the camera and the rest to a nearby PC. In hybrid operations, Bottlenose can be connected over standard ethernet using its GigE Vision API to popular industrial libraries like MVTec's HALCON® and Cognex's VisionPro®. It has on-camera processing for feature-point detection & matching, dense depth and DNNs. Its powerful ISP makes it ideal for mitigating complex lighting situations such as those in manufacturing facilities and factories.

These features allow integrators to use their existing programs on the PC while still being able to leverage the benefits of a smart camera.

Figure 2-4: Latest generation smart cameras have built in capability to process neural networks. This allows for unstructured and more flexible inspections. Obtained from [8] [9] [10].

2.1 Strengths of Edge Computing

Low latency: One strength of edge computing in industrial machine vision is its low latency [11]. This is because some camera data can be stored and processed on the device or a nearby PC, rather than being sent to a remote data center [12].

Speed & internet connectivity: Edge computing is fast. Since the edge device is near the source of data, data can be sent back and forth quickly. Edge computing doesn't require uninterrupted internet to work all the time.

Control & ownership: Companies that adopt edge processing and implement edge computing devices have control over that critical infrastructure. Since the companies own the edge devices, this allows them to modify the devices for their particular conditions. These modifications could include custom enclosures and backup power sources.

Scalability: When the camera can compute at the edge, the system is inherently scalable, because as users add more cameras they are, by default, adding the required computational resources. They are no longer restricted by the computer resources of the central processor.

2.2 Limitations of Edge Computing

Processing power: Computation resources deployed in edge processors are typically lower than in a data center, so the same algorithms can take longer to execute. Similarly, constrained storage and memory inherent in edge computing also affects processing speed.

Power consumption: DNN-based algorithms require intense computational power; this results in higher power consumption and heat dissipation. This is especially true when GPUs and FPGAs are used. Cameras that include on-board vision processing engines using ASIC designs have lower power consumption while still excelling in Tera Operations Per Second (TOPS) per watt of energy consumed [12].

Bandwidth limitations: Regardless of the architecture used to connect a camera to a CPU, the camera will always have bandwidth limitations. While this limitation is orders of magnitude better than sending data to be processed remotely, it does pose its own challenges. For example, a smart camera with a 1Gbps link has limited bandwidth for loading new models for AI, given that each model can be a couple of hundred MBs.

Compression: DNN models usually need compression techniques like pruning and quantization to run on a resource-constrained edge device. This has hindered adoption as some accuracy loss is to be expected. Today, high-end smart cameras can run a full floating point uncompressed neural network with 50+ million parameters directly on the camera, as with Bottlenose.

3 Cloud Computing In Industrial Machine Vision

Remote data centers are well-suited for training machine learning models, given their large storage and compute capabilities. Personal computers are not ideal for this task [13].

3.1 Strengths of Cloud Computing

Scalability: Cloud computing capabilities are infinitely scalable. Users are not limited by physical space, compute, power consumption or memory. Cloud computing capabilities allow users to add more compute or memory resources with a single click. Users are limited only by their operating-expense budgets.

Capex vs opex: Capital expenditure (capex) is more suitable for edge-computing investments, whereas operating expenditure (opex) is more suitable for cloud-based infrastructure. Cloud companies offer storage, elastic compute instances and enterprise applications [14] via pay-as-you-go models, i.e. opex. This is easier for teams seeking approval from senior leadership. Opex also doesn’t affect fiscal reports–keeping the shareholders happy.

Versatility and diversity: Cloud computing offers remote data center versatility and diversity, through nearly unlimited types of CPUs and GPUs available at the click of a button, as well as software available online. In this case, users are not only buying the app from marketplaces, but also renting the computer on which the app will run.

Improved access: Remote data centers allow for improved access. Since they are remote they let a user access information from anywhere, anytime they want [15].

Global workforce: Given that cloud computing capabilities allow updates and services to be implemented 24/7 companies can make full use of their global workforce [15].

Access to advanced software services: Cloud computing capabilities allow quick access to current versions of software [15].

3.2 LIMITATIONS OF CLOUD COMPUTING

Latency: Cyber-physical systems (CPS) rely on data from sensors to perform computationally-intensive tasks including decision making, learning and prediction, and data analytics [16]. Industries like smart traffic, AgTech, and manufacturing have communication latency requirements that can't be subject to delays. In the time it takes to relay data to the data center for processing and back, critical decision-making can be delayed. Even having data travel to a PC next to the sensor can cause enough latency to affect the system negatively, especially in manufacturing facilities where producing larger quantities every minute is of the utmost importance. Ultimately, latency can lead to significant losses.

Cybersecurity & privacy risks: For a cloud system to be considered secure it must achieve confidentiality, integrity and availability of outsourced data [17]. For data to be processed in the data center it must be encrypted and sent to cloud-based repositories. Saving consumer information from multiple consumers remotely could compromise the confidentiality of data [18]. Similarly, the mystery surrounding how cloud processing works, can put data owners at risk. Often, data owners lack the technical know-how to understand their data's physical location [18].

Legal complexity: The cloud comprises both hardware and software. The hardware portion of the cloud is subject to ordinary property rules; software falls into the intrinsically complicated framework of intellectual property rules [19]. The international nature of cloud computing, with data warehouses storing data from users in different countries with different legal frameworks, further complicates these legal issues. Copyright protection, database protection, data sovereignty, patents, and trademarks all need to be considered when evaluating the legal ownership of data, programs and hardware associated with cloud computing [19].

Other factors: It is estimated [18] that by 2025 nearly 50 billion IoT devices will be in circulation, which will weigh heavily on cloud service providers to ensure fast and secure access to data. Cloud computing investments often suffer from pushback from CFOs since they have a high probability of increasing operating expenses [20]. Using a cloud computing system is typically subject to subscription fees paid to the cloud service provider(s). Moreover, skipped or missed maintenance fees have the potential to lead to missed updates (new features, bug fixes) [21].

4 DECIDING WHERE THE EDGE LIES IN INDUSTRIAL MACHINE VISION

Once it's clear that models are trained in the cloud and vision processing will take place at the edge, designers must now decide where this "edge" lies. There are three options:

4.1 INDUSTRIAL PC

A large variety of cameras can be integrated with an IPC and they cost less than smart cameras shown in Figure 4-1. Most cameras either follow the GigE Vision® standard or the USB3 Vision standard, and can be swapped out. However, the lower cost of the cameras is offset by the high price of IPCs capable of vision processing. These IPCs are also large, have high power consumption and dissipate a substantial amount of heat, making them difficult to integrate in a factory setting.

Figure 4-1: Camera + GPU enabled IPC via GigE Vision or USB Vision. Obtained from [22] [23] [24].

This camera + IPC architecture would eventually talk to a PLC via an industrial bus like EtherNet/IP™, Profinet® or EtherCAT® to communicate the detection information. The PLC can then react and control actuators.

4.2 Smart Camera

A smart camera connected directly to industrial equipment like actuators, warning lights or PLCs has many benefits. They are small compared to the size of an IPC and typically don't generate as much heat. They are self-contained and therefore easy to set up. A couple of the more popular smart cameras are shown in Figure 4-2.

However, there isn't a huge variety of smart cameras to choose from. They can be expensive and require ongoing licensing. Programming method for smart cameras is not standardized across vendors and so cannot be switched out quickly or cost effectively. Integrators and factories typically buy into a certain vendor's product line.

Figure 4-2: Proprietary smart cameras. Obtained from [5] [7].

4.3 Smart Camera + Industrial PC

Recent updates to the GigE Vision standard have now included the ability to send data from smart cameras to IPCs. This approach would allow integrators and factories to use smart cameras without being tied into any particular vendor. In this mode of joint sensing, the majority of the processing is happening on the smart camera, which allows for factories to use smaller IPCs.

Smaller IPCs have many benefits including lower cost, better ruggedization and less heat dissipation than its larger counterparts making them suitable for factory environments.

There is a limited variety of smart cameras that follow the GigE Vision industrial standard, but this number is growing. One such example is shown in Figure 4-3, where multiple Bottlenose cameras can be connected to a $500 fanless IPC. Smart cameras with GigE Vision can still be swapped with regular GigE Vision cameras if the IPC is upsized. This would provide the integrators and manufacturers some assurance for software longevity.

Figure 4-3: Smart camera + compact IPC via GigE Vision. Obtained from [25].

4.4 CHOOSING THE RIGHT SOLUTION FOR YOUR APPLICATION

Identifying the application and scenario are key factors when it comes to deciding which solution is most appropriate for a given use case. In small factories where only a few cameras are required for vision, a good solution is likely the smart camera with light edge computing capabilities. However, this changes when we consider large factories in need of hundreds of cameras. In this scenario, the regular camera using heavy edge computing and remote data center could be a better choice because it:

- reduces the capex,
- simplifies maintenance and software upgrades,
- suits multi-source big data analysis and
- increases scalability.

5 CONSIDERATIONS FOR MACHINE VISION USING AI

When using artificial intelligence (AI) for machine vision, model classification services and model-training services operate in different time frames, so it is important to consider where these services will be located. By considering the strengths and limitations of edge and cloud computing with respect to both machine vision and AI, optimal placement of these AI services can be determined. Figure 5-1 highlights the use of cloud computing for uploading historical data, while emphasizing that edge processing should be used for real-time applications. Because images can be transmitted from the image capture equipment to edge processing over local ethernet, real-time vision analysis (using classification construction models) can be performed, while historical vision data can be sent to cloud computing services for more compute-intensive image identification and prediction modelling. These updated models can then be fed back to the edge to improve the real-time vision analysis.

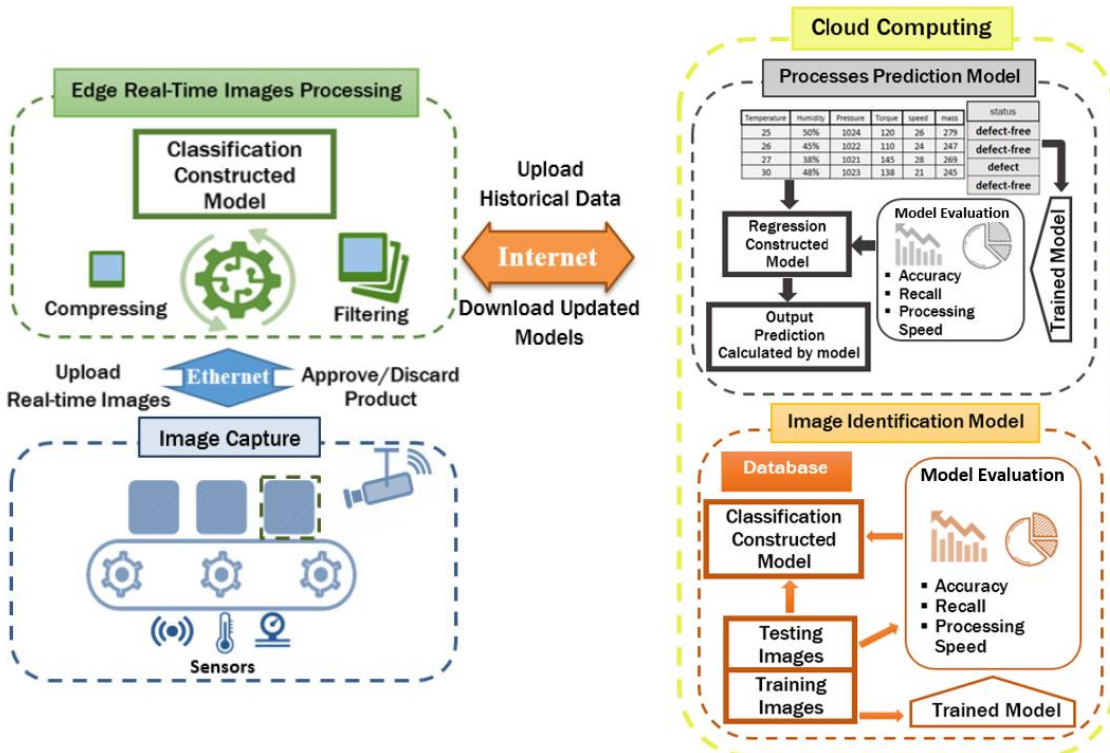

Figure 5-1: A visual representation of an intelligent machine vision model for defective product inspection. Obtained from [4].

6 Conclusion

Companies using industrial machine vision should adopt the flexibility and scalability offered by cloud computing for training their machine learning models. Edge computing should then be used as much as possible for the real-time industrial tasks.

This paper has outlined some considerations for where computation should be performed.

Bibliography

[1] M. Graves and B. G. Batchelor, “Machine vision for the inspection of natural products,” p. 471, 2003, Accessed: May 03, 2022. [Online]. Available: https://books.google.com/books/about/Machine_Vision_for_the_Inspection_of_Nat.html?id=PXwz4MDCkYsC

[2] “By any other name | Vision Systems Design.” https://www.vision-systems.com/boards-software/article/16738540/by-any-other-name (accessed May 03, 2022).

[3] R. Silva *et al.*, "Machine Vision Systems for Industrial Quality Control Inspections To cite this version : HAL Id : hal-02075633 Inspections," pp. 0–10, 2019.

[4] T. Benbarrad, S. B. Kenitar, and M. Arioua, "Intelligent machine vision model for defective product inspection based on machine learning," *2020 Int. Symp. Adv. Electr. Commun. Technol. ISAECT 2020*, 2020, doi: 10.1109/ISAECT50560.2020.9523643.

[5] V. Alonso, A. Dacal-Nieto, L. Barreto, A. Amaral, and E. Rivero, "Industry 4.0 implications in machine vision metrology: An overview," *Procedia Manuf.*, vol. 41, pp. 359–366, 2019, doi: 10.1016/j.promfg.2019.09.020.

[6] Wikimedia Commons contributors, "File:AutovisionIIatRDT.jpg," *Wikimedia Commons, the free media repository*, 2020.

[7] "What Are NI Vision Systems? - NI." https://www.ni.com/en-ca/innovations/white-papers/13/what-are-ni-vision-systems-.html (accessed Oct. 15, 2022).

[8] "Cognex - In-sight D900 Vision System." https://www.cognex.com/en-ca/products/deep-learning/in-sight-d900%0A (accessed Jul. 12, 2022).

[9] D. Edwards, "Robotics and Automation News," *SICK launches compact 2D vision camera with deep learning onboard*, 2021. https://roboticsandautomationnews.com/2021/04/16/sick-launches-compact-2d-vision-camera-with-deep-learning-onboard/42415/%0A (accessed Jul. 12, 2022).

[10] "Keyence." https://wwwd601d2yq4c.cdn.e2enetworks.net/content/products/1597923155Keyence_IV2.jpg (accessed Jul. 12, 2022).

[11] D. J. Deng, S. Y. Lien, C. C. Lin, S. C. Hung, and W. B. Chen, "Latency Control in Software-Defined Mobile-Edge Vehicular Networking," *IEEE Commun. Mag.*, vol. 55, no. 8, pp. 87–93, 2017, doi: 10.1109/MCOM.2017.1601165.

[12] F. Tao, J. Cheng, and Q. Qi, "IIHub: An industrial internet-of-things hub toward smart manufacturing based on cyber-physical system," *IEEE Trans. Ind. Informatics*, vol. 14, no. 5, pp. 2271–2280, May 2018, doi: 10.1109/TII.2017.2759178.

[13] J. Chen and X. Ran, "Deep Learning With Edge Computing: A Review," *Proc. IEEE*, pp. 1–19, 2019, doi: 10.1109/JPROC.2019.2921977.

[14] C. H. Choi, C. Lee, J. J. Lee, and K. Lee, "Understanding the Deployment Cost of Cloud Computing Services for the Higher Education Institutions," *ICTC 2019 - 10th Int. Conf. ICT Converg. ICT Converg. Lead. Auton. Futur.*, no. December, pp. 438–443, 2019, doi: 10.1109/ICTC46691.2019.8939863.

[15] Professor Gabriel Kabanda, "Performance of Machine Learning and Big Data Analytics Paradigms in Cybersecurity and Cloud Computing Platforms," vol. 21, no. 2, 2021.

[16] L. Bulej *et al.*, "Managing Latency in Edge-Cloud Environment," 2020, [Online]. Available:

http://arxiv.org/abs/2011.11450

[17] J. Domingo-Ferrer, O. Farràs, J. Ribes-González, and D. Sánchez, "Privacy-preserving cloud computing on sensitive data: A survey of methods, products and challenges," *Comput. Commun.*, vol. 140–141, no. March 2019, pp. 38–60, 2019, doi: 10.1016/j.comcom.2019.04.011.

[18] M. Mohammed Sadeeq, N. M. Abdulkareem, S. R. M. Zeebaree, D. Mikaeel Ahmed, A. Saifullah Sami, and R. R. Zebari, "IoT and Cloud Computing Issues, Challenges and Opportunities: A Review," *Qubahan Acad. J.*, vol. 1, no. 2, pp. 1–7, 2021, doi: 10.48161/qaj.v1n2a36.

[19] C. Bartolini, C. Santos, and C. Ullrich, "Property and the cloud," *Comput. Law Secur. Rev.*, vol. 34, no. 2, pp. 358–390, 2018, doi: 10.1016/j.clsr.2017.10.005.

[20] "Financial Impact of Moving to Cloud Computing," *Journal of Applied Business and Economics*, vol. 23, no. 1. 2021. doi: 10.33423/jabe.v23i1.4065.

[21] C. Fisher, "Cloud versus On-Premise Computing," *Am. J. Ind. Bus. Manag.*, vol. 08, no. 09, pp. 1991–2006, 2018, doi: 10.4236/ajibm.2018.89133.

[22] "Neousys New Edge AI GPU Computing Platform with NVIDIA® RTX A6000/ A4500," 2022. https://www.neousys-tech.com/en/news/press-room/2022/210-nuvo-8108gc-qd-edge-ai-gpu-computing-platform-with-nvidia-rtx-a6000-a4500 (accessed Jul. 22, 2022).

[23] "Sentech 2.35MP Monochrome USB3 Vision CMOS Camera with Global Shutter." https://www.bhphotovideo.com/c/product/1316985-REG/sentech_stc_mbs231u3v_2_35mp_black_white_usb3_vision.html (accessed Jul. 12, 2022).

[24] "Flea3 GigE." https://www.flir.ca/support/products/flea3-gige/%0A (accessed Jul. 12, 2022).

[25] OnLogic, "OnLogic Helix 310 Intel Elkhart Lake Industrial Edge Computer," 2022. https://www.onlogic.com/hx310/ (accessed Oct. 17, 2022).

AUTHORS & LEGAL NOTICE

Authors: The following persons contributed substantial written content to this document: Kendra Serbinski (Labforge Inc.), Yassir Rizwan (Labforge Inc.), Eugene Chang (Toshiba America Electronic Components, Inc.), Daniel Young (Toshiba America, Inc.).

Technical Editor: Stephen Mellor (IIC CTO) oversaw the process of organizing the contributions of the above Authors and Contributors into an integrated document.