Solution Overview

Intel® SceneScape
Industry 4.0

Edge-Based Multimodal Scene Intelligence for Digital Manufacturing

Intel and a large communication service provider (CoSP) deployed a demonstration architecture that simulates visuospatial understanding of complex scenes in a simulated industrial setting. Examples of use cases include asset tracking and automatic safety shutdown for industrial equipment. The proof of concept is installed at a Chicago venue operated by MxD (Manufacturing x Digital), a cross-industry institute for the development of transformative technology solutions for digital manufacturing.

5G networking improves machine-to-machine communication for implementations such as the industrial internet of things (IIoT). The high throughput, low latency, and advanced security of 5G make it possible to derive value from the large-scale data volumes generated by digital manufacturing. We now have a new generation of industrial solutions that draw on deep learning to drive insights and predictions from data streams. These solutions operate in nearly real-time and reduce backhaul costs by consuming and drawing intelligence from data at the network edge.

Intel® SceneScape is a platform for edge-resident multimodal scene intelligence that ingests sensor data from various types of sensors, aggregates it, and uses deep learning inference to understand the scene. Each data input enriches the whole scene, with Intel SceneScape applying each input as context to all the others. It maps the outputs of analytics based on these inputs to create a composite view tailored to specific application needs in the environment and maintains that view consistently over time to create a spatial-temporal graph that can be used to track, interpret, and predict the movements and behaviors of people and objects in the scene. High-level capabilities of Intel SceneScape include the following:

- **Scene and sensor management** maintains a scene map based on static "background" (static or slow-changing elements) and location, movement, and characteristics of various sensors such as cameras, microphones, and other sensors.
- **Real-time data fusion** combines inputs from all data inputs, across modalities, and transforms them into a form suitable for inclusion in the overall scene map.
- **Analytics** detect people, vehicles, and objects, interpreting meaningful events based on their movements and interactions, using capabilities such as geofencing, dwell time interpretation, and anomaly detection.
- **Scene APIs** provide data endpoints for scene data and interfaces for third party analytics and other applications, as well as maintaining trust relationships with users, devices, and software.
- **User interface capabilities** provide standards and connectivity across elements such as UIs, web resources, augmented/visual reality platforms, and gaming engines.

Assembling inputs from multiple diverse sensors, Intel SceneScape produces situational understanding of a scene in much the same way that human cognition does based on various sensory inputs. For example, the significance of smelling smoke depends on context provided by other senses as well as relevant pre-learned information. Thus, a person who smells smoke might seek to understand its relevance by looking for visual cues and understand its significance differently based on seeing a campfire versus a burning house. Likewise, Intel SceneScape combines various inputs from a physical scene, such as video images, sound, and temperature, interpreting them based on deep learning models that provide knowledge about potential circumstances in the scene and their significance.

Intel® SceneScape Demonstration Architecture

Intel and the CoSP implemented the demonstration architecture illustrated in Figure 1 as a functional proof of concept in a facility operated by MxD, a center for digital manufacturing excellence in Chicago. The institute performs research and proofs of concept for innovation around digitization and digital transformation for Industry 4.0 and beyond, in conjunction with tech leaders as well as academic centers and the defense industry.

The demonstration architecture is a flexible model using open components that can be adapted to varying needs, at any scale, which are described in more detail in the "Technology Building Blocks" section, below. Software elements of the demonstration architecture include the following:

- **The CoSP's multi-access edge computing (MEC) platform** provides high-throughput, low-latency operation with Intel SceneScape.
- **Intel® Smart Edge** is a commercial MEC software offering; the CoSP's MEC uses Intel Smart Edge components, which demonstrates that the platform can be integrated with other ecosystem solutions.
- **Intel® Distribution of OpenVINO™ toolkit** provides a platform to build, run, and optimize deep learning inference that is optimized for performance on Intel® processors and accelerators.
- **Intel SceneScape** is a platform that abstracts inputs from various sensors to create a semantic model in real time that supports software-based situational awareness.

The hardware infrastructure for the demonstration architecture is based on a single server equipped with accelerators for visual intelligence-related processing. The Dell PowerEdge R740 Server is a robust system well suited to deployment for demanding functions at the edge, powered by two Intel® Xeon® Scalable processors and Intel® Ethernet cards. In a real-world deployment, the infrastructure could be scaled to systems distributed to multiple edge locations.

Similar to a real-world implementation, the demonstration architecture provides a cyber-physical environment capable of gathering data using sensors and applying that data using controllers and actuators. The IoT endpoints enabled in the architecture include an array of four RGB cameras and a spare to capture visual data about the scene and a Mecademic Meca500 six-axis industrial robot arm to represent manufacturing equipment that can be controlled by applications.

These components are interconnected using a managed switch and 5G gateway, which handles communication between the IoT endpoints and the rest of the architecture. That link is through a 39 GHz millimeter-wave private 5G network, which provides secure low latency, high throughput wireless connectivity in factories and other large-scale facilities. This model eliminates the need for physical cabling, eliminating the high costs and inflexibility of hard-wired solutions and enhancing suitability for provisioning mobile equipment in real-world implementations.

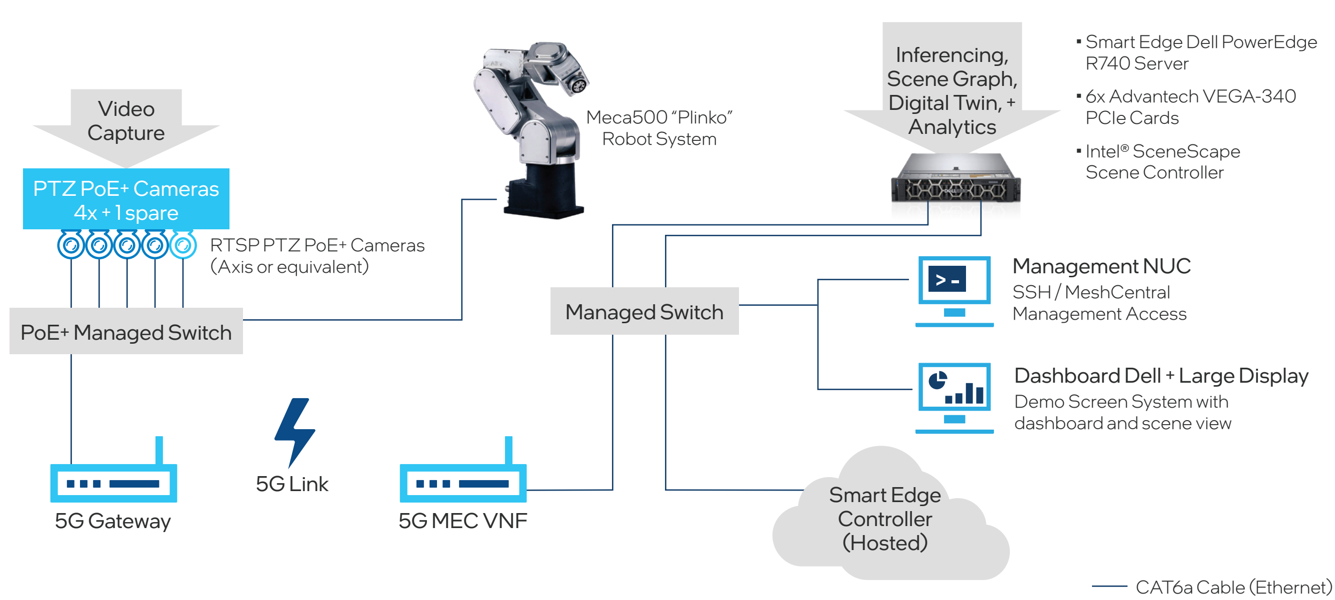

Figure 1. Topology for demonstration use cases.

The CoSP MEC manages the edge compute, including application provisioning and management, with secure orchestration for containerized workloads. Traditional implementations would have passed signals out of the building, out over the public network to the cloud or network core for processing, and then back into the building, with round-trip latency that would make it unsuitable for real-time digital usages such as digital safety and control for Industry 4.0.

The hosted Intel Smart Edge controller provides edge computing functionality that complements the CoSP MEC, such as hosting the Intel SceneScape Scene Controller that generates scene intelligence using computations performed on the PowerEdge R740 server. This proof of concept demonstrates successfully deploying components of both MECs together for combined operation.

Combining 2D visual inputs from multiple visual sensors, the solution calculates a representation of 3D space over time, creating a 4D conception of the scene. By maintaining a consistent representation of the 4D scene, the solution approximates spatial awareness to enable the use cases described below. The flexibility of this architecture is demonstrated by the ability to apply it to multiple use cases by changing only the algorithm. Thus, for example, a system built for visual identification of assets for an inventory system could be adapted in software to verify that personnel are wearing required safety equipment. The architecture is scalable to scenarios requiring thousands of sensors, as well as working with virtually any deep learning model, sensor, industrial controller, or data source.

Scene Intelligence Use Case 1: Industrial Safety Robot Shutdown

Visual intelligence is applicable to many industrial safety implementations, such as to monitor the position of human operators and to detect danger conditions if those people enter a hazardous area. Like the asset tracking use case, the industrial safety use case for this demonstration relies on existing and readily available deep learning models—in this case, algorithms to detect humans within complex scenes.

The Meca500 industrial robot arm is used with a fixed-position installation to represent factory equipment in this demonstration use case. To represent a relatively simple industrial process, the robot uses software developed by Intel engineers that enables it to play Plinko. Mapping the simulated factory floor that surrounds the robot, Intel SceneScape monitors the movement of humans in the area and automatically shuts the robot down if a human enters the designated "danger zone."

Plinko: The Origin Story

Debuting on the gameshow *The Price is Right* in 1983, Plinko features a contestant dropping flat, disk-shaped chips into a large vertical gameboard. Bouncing among a field of pegs as it drops through, the chip's random horizontal motion causes it to fall into one of several receptacles at the bottom of the board. Based on which receptacle the chip falls into, the contestant receives a corresponding prize.

As a simulated safety application, real-time operation is critical to this use case, as it is in many industrial use cases, such as process automation and control, which may be based on real-time operating systems (RTOSs) and other low-latency elements. In this use case, edge computing eliminates the transport latency that would be introduced by remote processing, so stimulus and response (danger condition and shutdown) can be nearly simultaneous. Data locality and 3GPP-compliant end-to-end security helps protect the integrity of the solution against cyber attacks or other tampering.

Scene Intelligence Use Case 2: Asset Tracking Using 2D Barcodes

Asset tracking is a fundamental use case in many industrial visual intelligence solutions, such as mapping the locations of capital equipment, raw materials, and finished goods. Deep learning models to visually identify common scene elements such as the presence of objects, humans, or animals are readily available, but industrial applications often need to identify specific individuals or objects, which requires new models to be manually developed and trained. Such development can be untenable for large-scale implementations where large numbers of changing elements need to be identified.

As an alternative to that approach, the asset tracking use case relies on AprilTags, which are 2D barcodes similar to QR tags, but simpler and designed for smaller data payloads. By printing an AprilTag and affixing it to a person or asset, operators have a simple, low-cost way of adding tracking functionality to any element of a scene. Because AprilTags are standardized and widely adopted, application logic can implement them based on existing deep learning models, without bespoke model development or training.

This approach enables tracking to be implemented with more flexibility, reduced cost, faster time to implementation, and more stability as new visual elements are added relative to the development of detection models from scratch. Visual tracking of assets is well suited to providing location data over time across large areas. It is more scalable than other approaches such as using passive RFID tags, which are more costly to deploy in large numbers.

In the asset-tracking use case demonstration, information is collected by all four cameras and ingested into Intel SceneScape's scene-intelligence engine to map the location of tagged people or objects into the scene. Intel SceneScape maintains a seamless spatial-temporal representation of the scene that spans the area of interest in its entirety, stitching together information from all available cameras so that movement of a tracked object from one camera's field of view to another is transparent to the tracking application. Tracking is accomplished using a fixed local Cartesian coordinate system bound by the covered area. It could be mapped to latitude, longitude, and elevation coordinates or another spatial system.

This use case also demonstrates the value of edge computing to optimize the amount of data sent over long-haul connections to a central point for processing. Because analytics are run locally on the R740 server, it is not necessary to send raw data long-distance over the wire, reducing bandwidth costs and making the use case viable at scale. At the same time, the private 5G network eliminates the need for cabling from cameras to edge-based analytics infrastructure.

Technology Building Blocks

The demonstration architecture uses standards-based technology building blocks to transform sensory data into scene intelligence by placing it into a dynamic scene graph that the solution maintains in real time. Each of those building blocks is discussed in more detail in the remainder of this section.

CoSP MEC

Using the private 5G network in the demonstration architecture, the CoSP MEC enables ultra-low-latency operation at the edge using high-speed wireless. The platform applies analytics to differentiate between data that should be kept local versus data that should be transmitted back to the cloud or the network core. This insight enables the CoSP MEC to enhance solution latency, data privacy, and long-haul bandwidth costs by consuming sensitive data close to where it is generated. It is explicitly built for simple, in-place upgradeability as technology roadmaps evolve, such as from 4G to 5G to 6G.

The MEC also provides ease of integration with both on-premise systems and cloud-hosted solution elements (such as the Intel Smart Edge controller in the demonstration architecture) on any of the popular public cloud infrastructures. Consumed as a virtualized network function, the MEC supports cloud-native operation in highly distributed edge networks, helping ensure future-readiness. It enables edge solutions that help derive value from 5G, integrating IoT and other machine-to-machine models into low-latency, high throughput business and industrial computing processes.

Intel® Smart Edge

Intel® Smart Edge is commercial software for the deployment and management of edge applications across any type of network. The stack includes components for various aspects of providing holistic, single-pane-of-glass visibility and control. It enables provisioning and maintenance of infrastructure with cloud-like agility and integration with popular tools, including the Intel Distribution of OpenVINO toolkit, as well as optimized performance across the full range of Intel processors and accelerators with a single codebase.

The demonstration architecture demonstrates the viability of deploying a subset of Intel Smart Edge components in conjunction with the CoSP MEC acting as the edge computing platform more generally. In this case, the CoSP MEC used elements from the Intel Smart Edge Open software stack for hosting Intel SceneScape, as well as for some support functions related to interfacing and setup. The standards-based architectures of both these toolsets support interoperation in an ecosystem-driven joint solution.

Intel® SceneScape

To emulate sensory awareness, Intel SceneScape ingests data from a variety of sensors, applies analytics to those streams to interpret them collectively, and creates a spatial-temporal scene graph as an abstracted representation of those data inputs as a function of time. In this representation, the data streams contextualize each other, enriching the aggregate, which is used to generate a digital twin model analogous to a "mind's eye" human understanding of the scene. Algorithms for semantic interpretation of this four-dimensional abstraction constitute multi-modal scene intelligence, such as detecting dangerous practices in a factory or tracking specific people and objects as they move around in a facility. The Intel SceneScape Scene Controller extends information to endpoints such as giving a robot access to real-time visual mapping of areas outside its direct scope of vision.

Intel SceneScape provides high performance from diverse sensor topologies, across processors and accelerators based on Intel architecture, to update spatial-temporal scene graphs continually for accuracy and maintain them consistently over time. To enable real-time operations including safety applications, real-time updates and low-latency access of scene graphs are typically enabled by placing Intel SceneScape at the network edge; for this demonstration architecture and most implementations, the platform is powered by Intel Smart Edge. Intel SceneScape provides open APIs that enable applications to access and control the scene graphs, powering solutions that understand what is happening in a scene, track people and objects through it, and predict what will happen next.

Intel® Distribution of OpenVINO™ Toolkit

The demonstration architecture uses the Intel Distribution of OpenVINO toolkit to streamline the processes of building, running, and optimizing deep learning models for human vision based on popular frameworks such as TensorFlow, Caffe, PyTorch, and others. Ingesting video stream data—RGB, infrared, or otherwise, software based on the toolkit applies inference-based analytics to interpret the visual data and plot it onto the scene graph as a continuous function of time.

The toolkit provides development tools, libraries, and pre-optimized kernels built specifically for convolutional neural networks (CNNs) associated with computer vision. In addition to helping ensure high performance on complex visual processing tasks, the toolkit advances a "write once, deploy anywhere" approach. With a simple configuration change, models can be adjusted to run inference on most combinations of processors and accelerators based on Intel architecture, including CPUs, GPUs, VPUs, FPGAs, and Intel® Gaussian & Neural Accelerators (Intel® GNAs).

Dell PowerEdge R740 Server

The PowerEdge R740 is a two-socket 2U rack-optimized server based on Intel® Xeon® Scalable processors. Built for expandability, the system supports up to eight PCI Express slots, 24 memory modules, and 16 local solid state drives or hard drives. It provides automated management capabilities out of the box, including customized reporting, automated discovery, and administrator accessibility using mobile devices. The R740 enhances reliability, availability, and serviceability (RAS) with fully redundant, hot plug power supplies and fans.

Cryptographically signed firmware packages and Secure Boot bolster the server's data-protection capabilities. Dell Supply Chain Security is a comprehensive set of practices that track and protect server components throughout their progress from the factory to the data center to help infrastructure architects guard against threats that include tampering and counterfeits. Hardware-based security capabilities built into the R740 include iDRAC9 Server Lockdown mode to mitigate malware and System Erase to wipe data from local storage media and memory.

Intel® Xeon® Scalable Processor

The compute engine that powers the PowerEdge R740 is the Intel® Xeon® Scalable processor, which offers high per-core performance, assisted by built-in acceleration technologies that include:

- **Intel® Advanced Vector Extensions 512 (Intel® AVX-512)** are processor instructions that enable ultra- wide 512-bit vector operations, for double the FLOPs per clock cycle compared to its predecessor.
- **Intel® Deep Learning Boost (Intel® DL Boost)** with Vector Neural Network Instructions (VNNI) eliminates unneeded precision in calculations for AI workloads, so they are completed more quickly.

Intel® Xeon® Scalable processors are built for use with other Intel® building blocks, including the Intel® Movidius™ Myriad™ X vision processing unit (VPU), which is used in this demonstration architecture.

Conclusion

Intel SceneScape emulates situational awareness by conjoining multiple sensory inputs to form an aggregate whole. The demonstration architecture built by a cross-industry collaboration shows the potential for industrial applications and beyond, in conjunction with 5G and edge computing. As an evolution of deep learning, Intel SceneScape brings computing one step closer to cognition.

More Information

Dell Technologies Edge Computing Solutions: delltechnologies.com/en-us/solutions/edge-computing/index.htm

Dell Technologies Telecommunications Solutions: delltechnologies.com/en-us/industry/telecom/index.htm

Intel® Distribution of OpenVINO™ Toolkit: intel.com/content/www/us/en/developer/tools/openvino-toolkit/overview.html

Intel Smart Edge: intel.com/content/www/us/en/edge-computing/smart-edge.html

MxD: mxdusa.org

Solution provided by:

intel + DELL Technologies + MxD

Features, SKUs, and frequencies are preliminary and subject to change.

Performance varies by use, configuration and other factors. Learn more at https://www.intel.com/PerformanceIndex.

Performance results are based on testing as of dates shown in configurations and may not reflect all publicly available updates. See configuration disclosure for configuration details. No product or component can be absolutely secure.

Intel is committed to respecting human rights and avoiding complicity in human rights abuses. See Intel's Global Human Rights Principles. Intel's products and software are intended only to be used in applications that do not cause or contribute to a violation of an internationally recognized human right.

Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy.

Your costs and results may vary.

Intel technologies may require enabled hardware, software, or service activation.

You may not use or facilitate the use of this document in connection with any infringement or other legal analysis concerning Intel products described herein. You agree to grant Intel a nonexclusive, royalty-free license to any patent claim thereafter drafted which includes subject matter disclosed herein.

The products described may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

© Intel Corporation. Intel, the Intel logo, and other Intel marks are trademarks of Intel Corporation or its subsidiaries. Other names and brands may be claimed as the property of others.

0522/DO/MESH/346435-001US