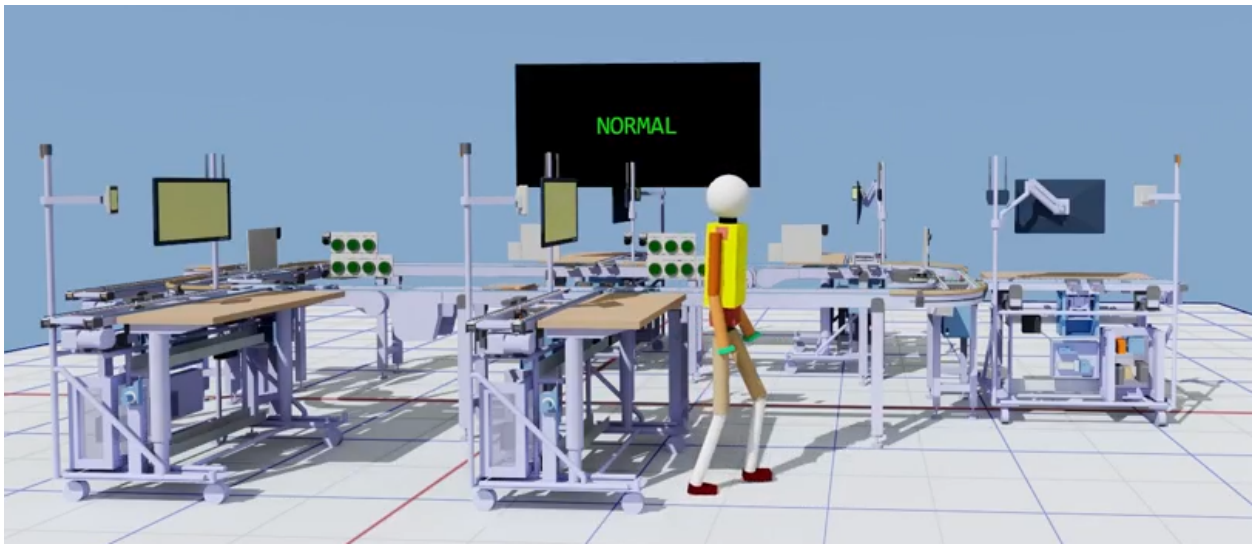
CHALMERS
UNIVERSITY OF TECHNOLOGY



A Human-in-the-Loop Digital Twin Architecture for Real-Time Safety-Control Simulations in Manufacturing Systems

A Human-centric approach in creating a digital twin using real-time motion capture Data

Master's thesis in Production Engineering

ALEN CHRISTOPHER & SURYA SAJEEV

DEPARTMENT OF INDUSTRIAL AND MATERIAL SCIENCE

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2026
www.chalmers.se

Master's thesis 2026

Creation of a Human-In-the-loop digital twin Architecture

A Human-centric approach in creating a digital twin using real-time motion capture Data

ALEN CHRISTOPHER
SURYA SAJEEV

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Industrial and Material Science
Division of Production Systems
Chalmers University of Technology
Gothenburg, Sweden 2026

Creation of a Human-In-the-loop digital twin Architecture
A Human-centric approach in creating a digital twin using real-time motion capture Data
ALEN CHRISTOPHER
SURYA SAJEEV

Supervisor: Pietro Lungaro, Occurence Technologies
Supervisor: Huizhong Cao, Industrial and Material Science
Examiner: Bjorn Johansson, Industrial and Material Science

Master's Thesis 2026
Department of Industrial and Material Science
Division of Production Systems
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Human Proxy integrated into the drone factory digital twin environment.

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2026

Creation of a Human-In-the-loop digital twin Architecture
A Human-centric approach in creating a digital twin using real-time motion capture Data
ALEN CHRISTOPHER
SURYA SAJEEV
Department of Industrial and Material Science
Chalmers University of Technology

Abstract

Digital twins are increasingly used to simulate and optimize manufacturing systems; however, current implementations are largely machine-centric, excluding human operators from virtual representation. This results in safety risks, reduced situational awareness, and limited support for human-machine collaboration central to Industry 5.0. This paper presents a human-in-the-loop digital twin architecture that integrates real-time human motion data into a simulated manufacturing environment for safety monitoring and human-centric production. The architecture spans three modular layers: a vision-based physical layer, a robust communication layer, and a simulation layer. Human motion is captured via the Occurrence multi-camera 3D pose fusion system, detecting 17 joints at 20 Hz, and streamed via MQTT to the Emulate3D simulation platform. A custom JSON parser and coordinate calibration pipeline transform incoming pose data into the simulation's coordinate space, with a dual-protocol fallback strategy ensuring connection reliability across varied network configurations. The live pose data is mapped into a hierarchical 14-joint human proxy model using a hybrid forward- and inverse-kinematics approach, enabling stable real-time replication of full-body motion. A spatially aware, dual-zone safety monitoring system computes real-time pelvis-to-conveyor distances using an Axis-Aligned Bounding Box (AABB) model, triggering warning alerts and latched automated emergency stops on zone violations. The architecture is demonstrated and validated at the drone assembly workstation in the SII lab at Chalmers University of Technology. Results from experiments confirm that all performance targets were met or exceeded, like the pose update rate of 20 Hz, end-to-end latency of 100 ms, etc. This demonstrates the potential of real-time human proxy integration with digital twin environments for adaptive safety monitoring and lays a scalable foundation for human-centric manufacturing.

Keywords: Digital Twin, Human-in-the-Loop Systems, Human-Centric Manufacturing, Real-Time Pose Estimation, Industrial Safety Monitoring, Human-Machine Collaboration, Industry 5.0, Cyber-Physical Systems.

Acknowledgements

We would like to express our sincere gratitude to our examiner, Professor Björn Johansson, and our supervisors, Huizhong Cao and Pietro Lungaro, for their continuous guidance, encouragement, and valuable feedback throughout this thesis project. Their expertise and support have been instrumental in shaping both the direction and quality of this work. We would also like to thank our technical and industrial partners, Occurrence Technologies, Rockwell Automation, and PTC, for providing the technologies, knowledge, and collaborative environment that made this research possible. Special thanks to the Digital Twin Innovation Testbed (DTIT) project for supporting this work and providing access to the research infrastructure used throughout the study. We would also like to acknowledge Simon Heyes, Abbe Ahmed, Niclas Jonasson, Marcus Gårdman, Sandra Jakšić, as well as Aeman Abdullah and Mohi Eddin Bilal for their technical assistance, insightful decisions, and continuous support during the development and the validation of the proposed architecture. Their contributions were invaluable to the completion of this thesis. Finally, we would like to thank all colleagues, researchers, and industry officials who provided feedback, encouragement and valuable discussions throughout this journey.

Alen Christopher and Surya Sajeev
Gothenburg, June 2026

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis listed in alphabetical order:

AABB	Axis-Aligned Bounding Box
AI	Artificial Intelligence
CPS	Cyber-Physical System
DT	Digital Twin
FK	Forward Kinematics
HiTL	Human-in-the-Loop
HiTL-CPS	Human-in-the-Loop Cyber-Physical System
HMI	Human-Machine Interaction
HCDT	Human-Centric Digital Twin
HTTP	Hypertext Transfer Protocol
IK	Inverse Kinematics
IoT	Internet of Things
ISO	International Organization for Standardization
JSON	JavaScript Object Notation
MQTT	Message Queuing Telemetry Transport
PLC	Programmable Logic Controller
QoS	Quality of Service
RQ	Research Question
SCADA	Supervisory Control and Data Acquisition
SII	Stena Industry Innovation Laboratory
TCP	Transmission Control Protocol
VR	Virtual Reality
WS	WebSocket
3D	Three-Dimensional

Nomenclature

Below is the nomenclature of indices, sets, parameters, and variables used throughout this thesis.

Indices

i	Index for tracked human joint
t	Time step / frame index
k	Joint hierarchy level index

Sets

J	Set of tracked human joints from the Occurrence system
J_p	Set of joints in the human proxy model
Z_s	Safe zone
Z_w	Warning zone
Z_d	Danger zone / emergency-stop zone

Parameters

f	Pose update frequency (Hz)
Δx	Calibration offset along the X-axis (m)
Δy	Calibration offset along the Y-axis (m)
Δz	Calibration offset along the Z-axis (m)
α	Rotation smoothing coefficient
d_w	Warning-zone threshold distance (m)
d_d	Danger-zone threshold distance (m)
N_j	Number of tracked joints

$y_{threshold}$	Ground-contact threshold height (m)
$v_{threshold}$	Foot velocity threshold for foot locking (m/frame)

Variables

$X_{occ}, Y_{occ}, Z_{occ}$	Joint coordinates in the Occurrence coordinate system
$X_{sim}, Y_{sim}, Z_{sim}$	Joint coordinates in the Emulate3D coordinate system
$\mathbf{p}_{joint}$	Position vector of a tracked joint
$\mathbf{p}_{pelvis}$	Pelvis position vector
$\vec{v}_{shoulder}$	Shoulder direction vector used for body orientation
$\hat{\mathbf{u}}$	Normalized upper limb vector
$\hat{\mathbf{v}}$	Normalized lower limb vector
d	Euclidean distance from pelvis to conveyor AABB
dx, dy, dz	Axis-wise distance components to AABB boundary
v_{foot}	Foot velocity between consecutive frames
θ	Joint rotation angle
θ_t	Filtered joint rotation at time step t
θ_{new}	Newly computed rotation angle before smoothing
θ_h	Hip joint rotation angle
θ_k	Knee joint rotation angle
θ_e	Elbow joint rotation angle
θ_s	Shoulder joint rotation angle
$pitch$	Pitch rotation computed from limb direction vector
yaw	Yaw rotation computed from limb direction vector

Contents

List of Figures

List of Tables

1

Introduction

1.1 Background

The increasing digitalization of manufacturing systems has positioned digital twins as a core enabling technology in Industry 4.0 and the emerging paradigm of Industry 5.0. Digital twins provide a dynamic virtual representation of a physical system, which is continuously updated via real-time data, enabling the ability to monitor, simulate, and optimize industrial processes [1, 2]. Within the manufacturing context, this monitor-simulate-optimize paradigm is applied to model machine behavior, production flows, and predictive analysis [3].

The adoption of digital twins in industrial settings has risen substantially over the past decade, driven by advances in sensing tech, industrial communication protocols, and simulation platforms. Researchers commonly use these tools to identify bottlenecks, optimize production layouts, and support operational planning [4]. In this context, the digital twin serves as a tool for supervisors and safety managers; it extends their situational awareness beyond that allowed by direct observation, providing a real-time virtual view of the interaction between humans and machines[5].

However, most of the implementations of digital twins currently are machine-centric. Human operators, who are central to most manufacturing processes, are rarely represented as active components in the digital environment [3]. Interactions between physical and virtual systems remain primarily screen-based and control-system-driven, making them less intuitive and poorly suited to capturing the real-time behavior of workers on the factory floor. This results in three key problems. Firstly, there is a lack of awareness of human activity, as worker presence is not dynamically reflected in the virtual environment. Second, the absence of human factors in system analysis reduces the applicability and realism of digital twin models [6]. Third, there is no mechanism for human intervention within digital twins, especially in safety-critical scenarios where the system responses must adapt in real-time to the presence and behavior of workers [7].

With the emergence of Industry 5.0, there has been a deliberate shift towards human-centric manufacturing, where collaboration between humans and machines plays a central role [6, 7]. Industry 5.0 extends the efficiency goals of Industry 4.0 by placing worker well-being, inclusivity, and societal value alongside productivity and technological progress. In this context, integrating human behavior into the digital twin

environment becomes critical to enable safer, more adaptive, and more responsive manufacturing systems [8, 6]. Human-centric digital twins have been proposed as an extension of existing implementations, broadening the scope of virtual representation to include human activities and behaviors [9].

Conventional approaches to capturing human behavior in manufacturing have relied on wearable-based motion capture systems, which, while effective for ergonomics studies, are intrusive, hygiene-sensitive, and impractical during normal production operations. Vision-based non-wearable motion capture offers a practical alternative by enabling real-time tracking of body joints and posture without any physical instrumentation on the worker [10]. However, translating raw pose data from such systems into a stable, automatically plausible digital representation within an industrial simulation platform remains a largely unsolved engineering challenge. This thesis addresses that challenge directly.

1.2 Research Gap

Existing research on digital twins has established strong foundations in architecture design, lifecycle modeling, and optimization of simulations. There are frameworks such as ISO 23247, which give standardized structures for the implementation of digital twins. Hence, there have been a lot of case studies of digital twins, mostly representing machines, robotic system and production lines.
However, when we consider research done on human digital systems and HITL CPS systems, major exploration was done in the conceptual approaches for human integration in digital environments. Human-centricity, assistance for operators, training, and enhancement of safety were mainly emphasized.
There were several limitations remaining, including:

- Digital twin implementations mostly prioritize machine behavior instead of representing human motion.
- Research on human digital twins mostly stays conceptual or focuses on visualization or scenarios regarding training.
- Structured methodologies regarding the transformation of real-time tracking data into an articulated digital proxy are very limited.
- Practical lab-scale implementations for non-wearable motion capture systems into a digital twin environment are also limited.
- Limited documentation regarding most technological and engineering challenges faced in this case, such as joint transformation mapping, animation robustness etc.

Hence, there is a gap between the conceptual ambition of Industry 5.0 and the practical realization of human digital twins.
This thesis addresses this gap by creating and demonstrating a structured methodology for real-time skeletal motion data integration into a simulation-based environment of the industrial digital twin (Drone Factory).

1.3 Problem Statement

Currently, digital twins lack the structured integration of real-time human motion data. As a result, digital twins are not able to represent human presence as a proper part of the digital twin environment and as a component of the digital twin. Without integrating this data, it is difficult to

- model scenarios regarding human-machine interaction (safety-critical)
- Simulate the collaborative work environments.
- Implement any proximity-based safety logic
- Develop human-aware Digital Twin applications

For enabling human-centric digital twin applications aligned with Industry 5.0 principles, there is a need for a systematic approach to capture non-wearable motion data, transform positional data into the articulated digital proxy, ensure the hierarchical motion representation is stable, and integrate the proxy into the existing digital twin environment.

1.3.1 Aim

The aim of this thesis is to develop, implement, and validate a structured framework for integrating real-time human motion data captured via non-wearables into the digital twin environment (Drone Factory), enabling human-centric digital twin applications.

1.3.2 Research Questions

To achieve the above-stated aim, the following research questions are addressed.
RQ1: How can non-wearable human motion capture data be transformed into a stable human proxy within a digital twin environment?
RQ2: How can hierarchical joint transformations and alignment of coordinates be implemented so that the motion can be represented consistently?
RQ3: How can the integrated human proxy be deployed within the Drone Factory's digital twin environment to demonstrate the human-centric functionality as based on the given use case?

1.4 Scope

The scope for this thesis includes:

- use of the non-wearable skeletal tracking system
- development of a human avatar (hierarchical) in Emulate3D.
- mapping of motion using Forward Kinematics.
- Calibration and stabilization of the joint behavior.
- prep for MQTT-based real-time streaming.
- deployment of avatar into the Drone factory digital twin environment
- Implementing a safety-oriented human-centric use case.

The emphasis is on feasibility, structured methodology for implementation, and a demonstrator-based validation.

2
Theoretical Background

This chapter presents the theoretical foundations underpinning the development of the proposed Human-in-the-loop digital twin architecture. It introduces the transition from Industry 4.0 to Industry 5.0, the concept and architecture of digital twins, human-centric digital twin systems, vision-based human motion capture technologies, kinematic modeling approaches, and communication architectures for real-time data integration. Together, these concepts establish the theoretical basis for integrating humans into manufacturing digital twins for safety-aware and human-centric production environments.

2.1 Digitalization and the Evolution Towards Industry 5.0

Digitalization refers to the integration of digital technologies into industrial processes to improve efficiency, flexibility, and decision-making capability. The emergence of Industry 4.0 introduced cyber-physical systems (CPS), enabling physical production systems to be monitored, analyzed, and controlled through real-time data exchange and digital representations [1]. Core enabling technologies of Industry 4.0—including the Internet of Things (IoT), digital twins, cloud computing, and advanced analytics—significantly improved operational visibility and efficiency in manufacturing systems [3].

However, Industry 4.0 remained predominantly focused on machine-centric optimization. Automated systems and robotic processes were prioritized, while the role of human operators was increasingly marginalized within the digital representation of production environments [7]. This created a fundamental disconnect: the physical manufacturing environment depends critically on human presence, judgment, and interaction, yet these were absent from the virtual models used for monitoring and decision support.

Industry 5.0 was introduced to address these limitations, shifting the focus from purely automated efficiency towards human-centric industrial systems [6]. Rather than replacing human involvement, Industry 5.0 advocates for meaningful collaboration between humans and machines, treating workers as central contributors whose capabilities should be supported through technology. Three core pillars define Industry 5.0: human-centricity, sustainability, and resilience [6]. For digital twin systems,

this transition implies that humans should no longer be treated as external actors but as active components of the cyber-physical system. Consequently, methods for capturing, representing and responding to human activity become essential for achieving the objectives of Industry 5.0 [9].

2.2 Digital Twins: Concept and Architecture

A digital twin is a dynamic virtual representation of a physical system that is continuously updated via real-time data, enabling the ability to monitor, simulate, and optimize industrial processes [1, 2]. The concept was originally introduced by Grieves and Vickers [2] in the context of product lifecycle management and has since been formalized and standardized through frameworks such as ISO 23247, which provides structured guidance for digital twin implementation in manufacturing [8].

A digital twin system typically consists of three core components [1]:

- **The physical system** — the real-world asset, process, or environment being represented.
- **The digital twin model** — the virtual representation that mirrors the state and behavior of the physical system.
- **The communication architecture** — the data exchange infrastructure that maintains synchronization between the physical and virtual systems.

Kritzinger et al. [3] provides a useful taxonomy that distinguishes between three levels of digital representation based on the degree of automated data exchange: the digital model (no automated data flow), the digital shadow (one-way automated data flow from physical to virtual), and the full digital twin (bidirectional automated data flow). This classification provides a useful framework to position the developed system in this thesis. The proposed architecture initially operates as a Digital shadow, where the real-time human motion data continuously updates the virtual environment, while safety responses generated in the Digital twin influence the operational state of the physical system.

Within the manufacturing context, the monitor–simulate–optimize paradigm of digital twins supports a range of key functions [11]:

- Real-time system monitoring and state visualization
- Process optimization and bottleneck identification
- Predictive maintenance and fault detection
- Simulation and testing of production scenarios
- Operator training and decision support

Despite these capabilities, existing implementations remain highly machine-centric, focusing on equipment behavior and system dynamics while excluding human operators from virtual representation [3]. Critical aspects such as human factors, ergonomics, and safety considerations therefore remain unrepresented in most current digital twin models [6].

2.3 Human-Centric Digital Twins and Human-in-the-Loop Systems

Traditional digital twin implementations treated human operators as external elements rather than active components of the system model. With the emergence of Industry 5.0, the need for human-centric digital twins that explicitly integrate human presence and behavior into the virtual representation has become a recognized research priority [9, 6].

Human-in-the-Loop Cyber-Physical Systems (HiTL-CPS) extend the traditional CPS model by integrating human operators as active system elements. This enables digital twins to represent human motion, actions, and interactions with equipment within the virtual environment in real time. The digital twin in this context serves as a cognitive augmentation tool for supervisors and safety managers, extending their situational awareness beyond what direct observation allows by providing a real-time virtual view of human-machine interaction [**?**]. This concept of cognitive augmentation is central to Industry 5.0, where technologies such as artificial intelligence, virtual reality, and sensor-based systems are used to support rather than replace human perception and decision-making [9].

Human-centric digital twins have been proposed as an extension of existing machine-centric implementations, broadening the scope of virtual representation to include human activities and behaviors [9]. Such systems enable several capabilities that are not possible in machine-only digital twins:

- **Real-time safety monitoring** — detecting human presence within defined zones and triggering automated responses.
- **Human-machine interaction analysis** — analyzing how workers interact with equipment in the virtual environment.
- **Ergonomic assessment** — evaluating worker postures and movement patterns to identify risk factors.
- **Operator training and simulation** — providing realistic virtual environments for training without exposing workers to physical risk.
- **Human-aware system optimization** — adapting production system behavior based on the detected presence and actions of workers.

Unlike traditional digital twins that focus primarily on machine behavior, human digital twins seek to represent the worker as an active and continously updated component of the virtual environment. This enables the digital twin to account for worker position, movement, and interaction with surrounding equipment, thereby supporting safer and more adaptive manufacturing options.

The integration of human presence into digital twins represents a fundamental step towards bridging the gap between physical communication and virtual interaction in manufacturing systems [8].

2.4 Vision-Based Human Motion Capture

Capturing human motion for digital twin integration requires technologies that can measure and reconstruct the positions and movements of the human body in real time. Two broad categories of motion capture exist: wearable-based and vision-based systems [10].

Wearable-based systems use inertial measurement units (IMUs), optical markers, or electromyographic sensors attached to the body to capture motion. While these systems can achieve high accuracy, they are intrusive, hygiene-sensitive, and impractical during normal production operations, as they interfere with work tasks and require extensive setup and calibration [10].

Vision-based, non-wearable systems use cameras and computer vision techniques to detect and track human body positions without any physical sensors on the worker. These systems typically extract the positions of key skeletal joints from camera input using deep learning-based pose estimation algorithms, producing a skeleton representation of the human body in real time [10]. Modern multi-camera systems extend this capability by fusing pose data from multiple viewpoints, improving robustness against occlusion and enabling three-dimensional joint localization with greater accuracy than single-camera setups.

The output of such systems is typically a set of joint coordinates in three-dimensional space, representing the positions of key anatomical landmarks such as shoulders, elbows, wrists, hips, knees, and ankles. In the system used in this thesis, the Occurrence multi-camera system produces 17 joint coordinates at a frequency of 20 Hz, providing a continuous stream of pose data that serves as the input to the digital proxy pipeline. Each joint coordinate is accompanied by a per-joint confidence score, which provides a basis for filtering low-quality detections before they are applied to the proxy model.

Despite their practical advantages, vision-based systems introduce challenges that must be addressed in the implementation. Occlusion — where parts of the body are hidden from the camera — can result in noisy or missing joint data. Fast or complex movements, particularly of the upper body, tend to produce less stable estimates than slower, more predictable lower body motions. Appropriate filtering and stabilization strategies are therefore necessary to ensure that the resulting proxy motion is plausible and stable enough for use in a real-time simulation environment.

The use of non-wearable motion capture is particularly attractive in manufacturing environments because it eliminates the need for workers to wear dedicated tracking devices, reducing operational disruption while enabling continuous monitoring of human activity.

2.5 Kinematic Modeling for Human Proxy Animation

Translating raw joint position data from a motion capture system into meaningful, anatomically plausible movement within a digital proxy model requires appropriate kinematic modeling techniques [12]. Two fundamental approaches exist: forward kinematics and inverse kinematics.

2.5.1 Forward Kinematics

Forward kinematics (FK) computes the position and orientation of each joint based on predefined hierarchical relationships and joint transformations applied from the root of the skeleton outwards. In FK, the pose of a child joint is determined by the cumulative transformations applied along the chain from the root joint [12]. This approach is well suited to real-time simulation because it is computationally efficient and does not require iterative solving. Given the joint positions provided by the motion capture system, FK can be used to compute the rotations required to orient each segment of the proxy model by calculating directional vectors between adjacent joints and decomposing these into rotation angles using trigonometric functions such as the arc tangent.

2.5.2 Inverse Kinematics

Inverse kinematics (IK) approaches the problem from the opposite direction: given a desired end-effector position (such as the position of the foot or hand), IK computes the joint configurations required to achieve that position [12]. The FABRIK algorithm (Forward and Backward Reaching Inverse Kinematics), proposed by Aristidou and Lasenby [12], is a widely used iterative IK solver that converges quickly and produces natural-looking results. While IK provides greater accuracy for end-effector positioning, it is generally more computationally demanding than FK and can be less stable under noisy input conditions.

2.5.3 Hybrid Kinematic Approach

In practice, neither pure FK nor pure IK is optimal for all joints in a real-time human proxy driven by noisy motion capture data. Pure FK applied to leg joints, for example, cannot guarantee stable ground contact, as the foot position is determined entirely by the chain of rotations from the pelvis downwards and may float above or penetrate the ground plane when input data is noisy. Pure IK, on the other hand, requires additional computation power that may not be available within the constraints of a 20 Hz scripted simulation environment.
The hybrid approach adopted in this thesis applies FK to upper-limb joints, where computational efficiency is prioritized, and IK-based constraints to lower-limb joints

to ensure stable ground contact and eliminate stance-phase jitter. This combination provides a practical balance between real-time performance and anatomical plausibility, as discussed in Chapter 4.

2.5.4 Joint Types and Rotation Computation

Human joints exhibit different degrees of freedom depending on their anatomical function. Ball joints — such as the hip, shoulder, and wrist — allow rotation about multiple axes and are modeled using two-axis rotation computed from the normalized direction vector of each limb segment. Hinge joints — such as the knee and elbow — allow rotation about a single axis and are modeled using a single-axis flexion angle computed from the dot product of adjacent normalized limb vectors [12]. This differentiation ensures that the proxy model produces physiologically plausible motion and suppresses anatomically impossible joint configurations that could otherwise arise from noisy input data.

2.6 Communication Architectures for Real-Time Digital Twin Integration

Enabling real-time integration of human motion data into a digital twin environmentrequires an efficient and scalable communication architecture between the sensing system and the simulation platform [1, 11]. In industrial cyber-physical systems, communication architectures typically consist of several functional layers:

- **Sensing and data acquisition layer** — physical sensors and cameras that capture real-world state data.
- **Communication layer** — protocols and middleware that transmit data between physical and virtual systems.
- **Processing layer** — components that parse, filter, and transform raw data into formats suitable for the simulation platform.
- **Digital twin platform layer** — the simulation environment that receives, processes, and visualizes the data.
- **User interface and feedback layer** — displays and alerts that communicate system state to operators and supervisors.

2.6.1 MQTT Protocol

Message Queuing Telemetry Transport (MQTT) is a lightweight publish-subscribe messaging protocol widely used in cyber-physical systems and IoT applications [1]. MQTT operates on a broker-based architecture, in which publishers send messages to central broker on named topics, and subscribers receive messages by subscribing to those topics. This decoupled architecture enables scalable, low-latency data distribution across distributed systems without requiring direct connections between producers and consumers of data.

MQTT's key characteristics make it well suited to real-time human motion streaming in digital twin applications:

- **Lightweight overhead**—MQTT has minimal protocol overhead compared to HTTP-based alternatives, making it suitable for high-frequency data streams such as 20 Hz pose data.
- **Publish-subscribe decoupling** — the sensing system and simulation platform do not need to be directly connected, improving modularity and scalability.
- **Quality of Service levels** — MQTT supports configurable delivery guarantees, enabling reliable transmission even under variable network conditions.
- **Dual transport support** — MQTT can operate over both TCP and WebSocket transports, providing fallback options when network or firewall configurations restrict one of the ports.

In this thesis, MQTT is used as the primary communication channel between the Occurrence pose estimation system and the Emulate3D simulation platform, with a dual-protocol fallback strategy implemented to ensure continuous data flow across different network configurations [1, 11].

2.7 Design Science as a Research Methodology

The development of integrated systems such as the one proposed in this thesis aligns with design science research methodology, which focuses on the creation and evaluation of artifacts to address practical problems [13, 14]. Design science emphasizes iterative development, evaluation, and refinement of solutions within real-world contexts, ensuring both theoretical grounding and practical contribution [13].

In the context of human-in-the-loop digital twins, design science supports the development of architectures that integrate sensing, communication, and simulation components into a coherent system [15]. The iterative nature of the methodology is reflected in the development process followed in this thesis, where the human proxy model, kinematic mapping, communication pipeline, and safety monitoring framework were each developed, tested, and refined based on empirical observations from testbed experiments at the SII Lab, Chalmers University of Technology.

3

Methods

3.1 Research Methodology

This thesis employs a design science methodology focusing on creating and evaluating artifacts to address practical problems [13, 14]. This approach emphasizes iterative development, evaluation, and refinement within real-world contexts, ensuring theoretical grounding as well as practical contribution [13]. This approach is perfect for the present work, as the central objective is the design, implementation, and validation of a functioning system artifact, the human-in-the loop digital twin, rather than the testing of a predefined hypothesis.

The research process was structured into six interconnected phases:

1. **Literature review and theoretical framing**—Theoretical Foundations that support this work is established and verified (Related to Industry 5.0, Cyber-physical systems, HiTL digital twin systems, kinematics and motion mapping, communication architectures and safety-control systems.)
2. **Infrastructure analysis and system understanding**—The Drone Factory Testbed at the SII-Lab is analyzed, including the Occurrence multi-camera motion capture system, MQTT communication architecture, and Rockwell's Emulate3D simulation platform. Mainly done for identification of technical constraints and requirements for integration.
3. **Human-in-the-Loop Digital Twin architecture design**—The layered system architecture consisting of the physical motion capture layer, the communication layer and the virtual layer (simulation software) is defined for integrating human motion into the digital twin environment.
4. **Human proxy development and motion mapping**—Hierarchical human proxy model is designed, and the hybrid kinematics-based motion mapping pipeline is implemented for real-time replication of human motion inside the digital twin.
5. **Communication and safety monitoring implementation**—MQTT-based streaming pipeline is developed, with a dual-protocol MQTT fallback implemented (TCP/WebSocket). The safety intelligence framework is also integrated, with real-time safety monitoring including warning and emergency stop logic.
6. **Validation and demonstrator-based testing**—Evaluation of the completely integrated architecture within the Drone Factory testbed via live motion replication tests, communication robustness testing, and safety zone validation scenarios.

These phases may seem sequential, but the research process was highly iterative. The information received during implementation and testing continuously affected earlier design decisions, especially regarding coordinate alignment, hierarchy design of the model, motion stabilization and safety zone logic. This iterative approach enabled continuous refinement of the system architecture and ensured technical feasibility, real-time performance and consistency across all components of the architecture.
The overall methodology aligns with the Design Science Research process proposed by Peffers et al. [14], where the identified problem—the absence of human integration in conventional digital twins—guided the development, demonstration, and evaluation of the proposed Human-in-the-Loop Digital Twin architecture.

3.2 Literature Review and Conceptual framing

The literature review was conducted with two initial objectives: establishing the conceptual foundation for the work and identifying the technical enablers of integration of humans into digital twin environments. The review covered the following topics:

- Industry 5.0 and human-centric manufacturing principles [6].
- Digital twin architectures and frameworks [1, 3, 8].
- Human-in-the-loop cyber-physical systems [9].
- Vision-based motion capture systems [10].
- Kinematic modeling approach for human proxy animation [12].
- IoT communication protocols, particularly MQTT [11, 1]
- Design Science Research Methodology [13, 15, 14].

The review established that while conceptual discussions of human-centric digital twist exist in the literature, structured methodologies for real-time integration of non-wearable motion capture data into industrial digital twin virtual environments are very limited. This directly shaped the design choices made in this thesis, especially the decision to develop a custom kinematic mapping pipeline and a purpose-built hierarchical proxy model rather than relying on existing avatar frameworks.
This review wasn't confined to the beginning of the thesis but also continued throughout development, especially when addressing modeling decisions such as hierarchy design, coordinate transformation strategy and motion stabilization approaches.

3.3 System Understanding and Infrastructure analysis

Before implementing, the existing infrastructure at the SII Lab, Chalmers university of Technology was analyzed to understand the constraints that may exist for human integration and technical possibilities. The system architecture consists mainly of the following components:

- **Occurrence's multi-camera system**– a non-wearable vision-based pose estimation system that extracts 17 joint coordinates at 20 Hz and publishes the data via MQTT.

- **MQTT broker** – the communication middleware, which facilitates the real-time data exchange between the motion capture system and the simulation platform.
- **Rockwell's Emulate3D 2025** – the industrial simulation platform used for modeling the Drone Factory digital twin and hosting the human proxy model.
- **Drone Factory Workstation setup** – the physical lab environment at SII lab, consisting of a drone assembly workstation with a conveyor system and different stations for assembly.

Key aspects given special attention during the infrastructure analysis included the coordinate system misalignment between the occurrence visualizer, modeled after the original lab coordinates, and the drone factory digital twin coordinate system in Emulate3D, data serialization formats; update frequencies; platform scripting capabilities and restrictions; and the network configurations, which affected the MQTT connectivity.

A notable constraint identified during this phase was that Emulate3D does not natively accept MQTT payloads in array format, and its scripting environment imposes restrictions on runtime library loading. This constraint shaped the custom communication solution described in Section 3.5.

3.3.1 System Architecture:

The overall architecture is structured into three functional layers that together form the human-in-the-loop digital twin pipeline, as illustrated in Figure **??**:

- **Physical Layer** – Handled via the Occurence multi-camera motion capture system that feed an AI model which produces 17 joint keypoints at 20 updates per second.
- **Communication Layer** – Bridge between Occurence system and the simulation environment in Emulate3D. It recieves the JSON pose data, parses it at high frequency, locks onto the right operator ID, transforms the coordinate system and delivers clean skeletal data to the simulation.
- **Virtual Layer** – Consists of our Simulation platform, the digital twin simulation environment where our Human proxy model lives, along with the visibility control logic, the motion stabilization algorithms and the complete safety intelligence system.

The seperation of concerns across these three layers supports modularity and scalability, consistent with established layered architectures in cyber-physical systems as well as digital twin systems [1, 11, 4]. Each layer is described in detail in the following sections.

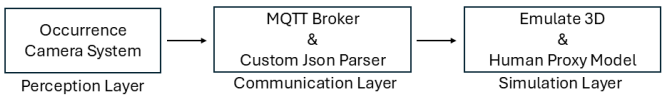


Figure 3.1: Layered architecture of the human-in-the-loop digital twin system

3.4 Physical Layer: Human Motion Capture

The physical layer consists of the real-world environment in which stable and high-frequency human motion is captured using the Occurrence multi-camera system. The system uses AI-based algorithms to detect and track human subjects, producing pose data for 17 joints at a frequency of 20 Hz. The 17 tracked joints follow a fixed ordering, as depicted in figure 3.2.

The sensing framework was designed around five key objectives:

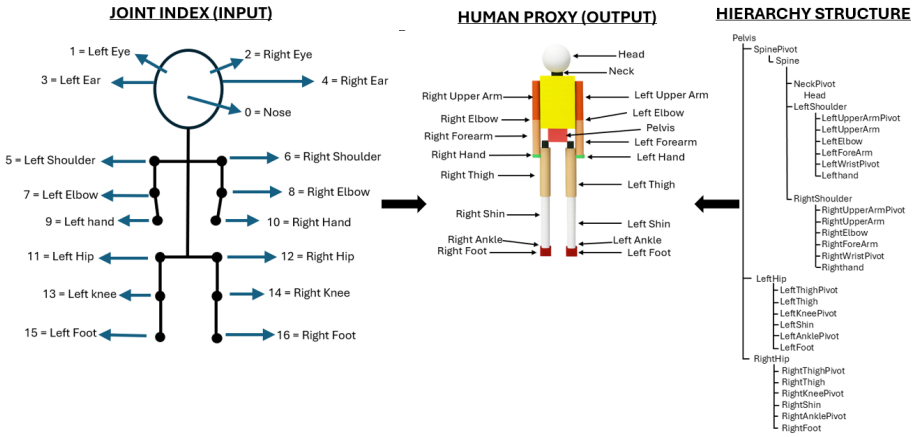


Figure 3.2: Joint Coordinate Mapping – Joint Index (left), Human Proxy (middle), and Hierarchical Structure (right)

- **Multi-view fusion**—synchronized camera fusion was used to improve spatial tracking accuracy and reduce occlusion-related tracking loss.
- **High-frequency pose estimation**—The system, as mentioned above, continuously tracks 17 joints at an update frequency of 20 Hz, giving smooth real-time sync with the digital twin environment.
- **Structured data streaming**—Skeletal joints were continuously streamed via MQTT using lightweight JSON payloads optimized for low-latency transmission. The coordinate Data is streamed continuously to the broker under the topic cm/{site_id}/fused for downstream consumption by the communication layer.
- **Spatial and ID filtering**—unique operator IDs and spatial windowing mechanisms were implemented to isolate individual workers inside predefined workstation zones. As it detects multiple persons, each person gets assigned unique ID numbers.
- **Probabilistic noise reduction**—Confidence-score thresholding is applied to reduce low-confidence detections and decrease sensor -induced jitter during rapid movements.

Therefore it serves as the primary perception system of the architecture, providing real-time spatial information about the human operator to the downstream communication and simulation layer.

3.5 Communication Layer Development

The communication layer facilitates real-time data transfer from the physical sensing system to the virtual simulation environment using the MQTT publish-subscribe protocol [1]. MQTT was selected for its lightweight architecture, low communication overhead, and suitability for high-frequency real-time data streaming across distributed cyber-physical systems [1, 11]. The layer encompasses four key components: runtime library integration, custom JSON parsing and ID locking, a dual-protocol fallback strategy and coordinate system alignment.

Figure 3.3: Communication Layer Pipeline.

3.5.1 Dynamic Runtime Library Integration.

Emulate3D does not natively accept array-format MQTT payloads, and its scripting environment imposes restrictions that prevent standard library imports at compile time. To address this, the MQTTnet v4.3.6 library was loaded dynamically at runtime using the `Assembly.LoadFrom()` method via .NET reflection. The MQTT client factory, client instance, connection options, and topic subscriptions are all constructed through runtime reflection using `Activator.CreateInstance()` and dynamic method invocation, bypassing compile-time dependency requirements:

```
var mqttAssembly = Assembly.LoadFrom(@"C:\...\MQTTnet.dll");
dynamic factory = Activator.CreateInstance(
    mqttAssembly.GetType("MQTTnet.MqttFactory"));
_mqttClient = factory.CreateMqttClient();
```

This approach enabled all MQTT operations – including connection management, topic subscription, and message reception – to be performed dynamically within the simulation scripting environment without requiring changes to the platform itself. The message handler is registered as an asynchronous delegate on the MQTT client's `ApplicationMessageReceivedAsync` event.

3.5.2 Custom JSON parser and ID locking

A custom JSON parser was implemented within the human proxy script to convert the incoming array-format payload into `Vector3` coordinate objects and map them to the 17 joints of the proxy model. Two pre-compiled regular expressions are used for high-frequency parsing: one pattern targets the `id_1` field to extract person identifiers, and a second triplet pattern extracts all `[x, y, z]` coordinate arrays from the targeted person block. Scientific notation and signed floating-point values are fully supported in the triplet pattern:

```
var pose3dMatch = Regex.Match(payload,
    @"pose_3d\s*:\s*\[(...)\",
    RegexOptions.Singleline);
var triplets = Regex.Matches(
    pose3dMatch.Groups[1].Value,
    @"\[\s*([+-]?\d+\.?\d*)\s*,\s*([+-]?\d+...)\s*\]");
```

The system locks onto the first detected person upon receiving data, recording that person's `id_1` value. This locked ID is maintained for the duration of the session, ensuring that only one worker drives the proxy at any time and preventing inadvertent switching between persons mid-session. If the locked target is absent for more than 30 consecutive frames, the lock is released and automatically re-acquired on the next detected person.

3.5.3 Dual-Protocol Fallback Strategy

To ensure communication robustness across different network configurations, a dual-protocol connection strategy was implemented. The system first attempts to establish a connection using MQTT over TCP (port 1883). If this fails due to network

restrictions or firewall configurations, it automatically falls back to MQTT over WebSocket (port 443). User credentials and port credentials are embedded in the connection logic to facilitate dynamic switching. This strategy ensures continuous data flow regardless of the network environment and was successfully validated across multiple network configurations during testing.

3.5.4 Coordinate system Alignment

The coordinate systems of the Occurrence camera system and the Emulate3D simulation platform are not identical and require explicit alignment. A physical calibration procedure was performed in which human subjects stood at known reference positions within the workspace. The real-time coordinates reported by the Occurrence system were compared to the corresponding positions in the Emulate3D scene, and the differences were used to compute calibration offsets Δx, Δy, and Δz. These offsets are applied to every incoming frame according to the transformation:

$$X_{sim} = X_{occ} + \Delta x \tag{3.1}$$

$$Y_{sim} = Y_{occ} + \Delta y \tag{3.2}$$

$$Z_{sim} = -Z_{occ} + \Delta z \tag{3.3}$$

The Z-axis is inverted ($Z_{sim} = -Z_{occ} + \Delta z$) due to a difference in axis orientation conventions between the two systems; without this inversion, the proxy model appears to face in the opposite direction to the physical subject. The calibrated offsets applied in the implementation are $\Delta x = 1.15\,\text{m}$, $\Delta y = 0.0\,\text{m}$, and $\Delta z = 0.005\,\text{m}$, with an additional empirical correction of $-0.04\,\text{m}$ applied to the pelvis Y-coordinate to compensate for the geometric offset between the hip midpoint and the physical pelvis center:

```
pelvis.RelativeLocation.X = vPelvis.X + offsetX; // 1.15f
pelvis.RelativeLocation.Y = vPelvis.Y - 0.04f;
pelvis.RelativeLocation.Z = -vPelvis.Z + offsetZ;
```

3.5.4.1 Resilient Communication Framework

During integration it was observed that different network environments handled MQTT communication differently. While some environments supported MQTT over TCP, others restricted direct TCP traffic and only allowed MQTT over WebSocket. To ensure communication robustness and deployment flexibility, a dual-protocol fallback mechanism was implemented. The system first attempts connection through MQTT TCP and automatically switches to MQTT WebSocket if the primary connection fails. This ensured continuous low-latency synchronization across varying network conditions and prevented runtime communication interruptions.

3.6 Virtual Layer: Human Proxy Model

The virtual layer consists of the Human Proxy framework, the kinematic motion stabilization framework, safety monitoring, camera zone detection and the real-time safety-monitoring architecture, all implemented within Rockwell's Emulate3D simulation platform using C# scripting. This layer represents the primary technical contribution of the work, translating raw skeletal coordinates into a stable, articulated virtual human that interacts meaningfully with the simulated manufacturing environment.

3.6.0.1 Proxy Model design

A custom human proxy model was designed and built from scratch in Emulate3D. The model uses basic geometric shapes—spheres, cylinders, and boxes—scaled to average human body dimensions. Rather than importing an existing avatar asset, the proxy was constructed from first principles to ensure full programmatic control over every articulation point. The model is controlled through 14 joint pivot points structured as a parent-child chain that mirrors the kinematic hierarchy of the human body. The pelvis serves as the root node of the hierarchy, with all other body segments defined as its children. The full hierarchy is structured as follows:

- Pelvis (root)
  - SpinePivot → Spine → NeckPivot → Head
  - LeftShoulder → LeftUpperArmPivot → LeftElbow → LeftForeArm → LeftWristPivot → LeftHand
  - RightShoulder → RightUpperArmPivot → RightElbow → RightForeArm → RightWristPivot → RightHand
  - LeftHip → LeftThighPivot → LeftKneePivot → LeftShin → LeftAnklePivot → LeftFoot
  - RightHip → RightThighPivot → RightKneePivot → RightShin → RightAnklePivot → RightFoot

The five head joint coordinates provided by the Occurrence system (eyes, ears, and nose) are averaged into a single centroid point, which is mapped to the head sphere of the proxy to replicate head motion. The 17 raw joints from the perception layer are therefore mapped into this 34-part hierarchical structure, with pivot nodes interspersed between geometry nodes to enable independent rotational control at each articulation point.

3.6.0.2 Kinematic mapping

Joint rotations are computed each frame using a hybrid forward kinematics and inverse kinematics approach. Forward kinematics (FK) is used for all limb segments by computing joint angles from the directional vectors between adjacent joints. An IK-inspired foot stabilization constraint is applied to the legs to ensure ground contact and eliminate stance-phase jitter. This hybrid approach was chosen deliberately: pure FK alone is insufficient because it cannot resolve the floating foot artifact caused by the absence of ground contact data in the ankle coordinates; pure IK, on the other hand, requires additional computational resources not compatible with a

20 Hz scripted simulation loop. The combination provides real-time plausible motion replication at the required update frequency [12].

3.6.0.3 Pelvis Position and Global Orientation

The pelvis position is computed each frame as the midpoint between the left and right hip coordinates:

$$\mathbf{p}_{pelvis} = \frac{1}{2}\left(\mathbf{p}_{LeftHip} + \mathbf{p}_{RightHip}\right) \tag{3.4}$$

The shoulder vector is used to derive the global yaw orientation of the body:

$$\vec{v}_{shoulder} = \mathbf{p}_{RightShoulder} - \mathbf{p}_{LeftShoulder} \tag{3.5}$$

The yaw angle is computed using the `atan2` function applied to the z and x components of the shoulder vector, with a 180° offset applied to align the proxy's forward direction with the physical object. A low-pass filter with smoothing coefficient $\alpha = 0.1$ is applied to suppress jitter during body rotation:

$$\theta_t = (1 - \alpha)\,\theta_{t-1} + \alpha\,\theta_{new} \tag{3.6}$$

A boundary correction prevents discontinuous jumps at the 0°/360° boundary by normalizing the angular difference $\Delta\theta$ into the range $[-180, 180]$ before accumulating it into the smoothed estimate. The spine direction is computed each frame as the vector from the pelvis midpoint to the shoulder midpoint, and a partial pitch and yaw (scaled to 0.6 and 0.3, respectively) are applied to the spine pivot to produce a natural torso lean without over-articulation.

3.6.1 Pelvis-Local Transformation Correction

A critical challenge encountered during development was unintended leg twisting when the body rotated. This arose because world-space joint vectors were being applied to joints that are defined relative to the pelvis local frame, causing leg orientation to behave incorrectly during global yaw rotation. This was resolved by transforming all world-space vectors into the local coordinate frame of the pelvis before computing joint rotations. The transformation rotates each input vector by the negated smoothed yaw angle $-\theta$:

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} \cos(-\theta) & 0 & -\sin(-\theta) \\ 0 & 1 & 0 \\ \sin(-\theta) & 0 & \cos(-\theta) \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{3.7}$$

This correction stabilized lower limb articulation, eliminated mirrored leg behavior, and removed rotational drift during turning. It is applied to all hip, shoulder, and wrist joint direction vectors prior to rotation computation.

3.6.2 Ball Joint Rotation

Hip, ankle, and wrist joints are modeled as ball joints with two degrees of rotational freedom. Pitch and yaw rotations are computed from the normalized direction vector of each limb segment:

$$\text{pitch} = \text{atan2}(d_z,\ -d_y) \tag{3.8}$$

$$\text{yaw} = \text{atan2}(d_x,\ -d_y) \tag{3.9}$$

Shoulder joints use a modified variant that clamps the Y-component to a minimum absolute value of 0.02 to avoid numerical instability near the vertical singularity and applies asymmetric scaling to the right shoulder's yaw axis (0.7) to better match observed motion in the physical environment. Wrist joints apply an additional 0.5 attenuation to both pitch and yaw to reduce over-articulation caused by noisy hand tracking data.

3.6.3 Hinge joint rotation

Knee and elbow joints are modeled as hinge joints with a single axis of rotation. The flexion angle is computed from the dot product of the normalized upper and lower limb vectors:

$$\theta = \cos^{-1}(\hat{\mathbf{u}} \cdot \hat{\mathbf{v}}) \tag{3.10}$$

A dead zone of 10° is applied to suppress small oscillations caused by sensor noise, so that angles below this threshold are set to zero. For knee joints, the computed angle is additionally clamped to a maximum of 175° to prevent hyperextension artifacts. The rotation axis for knee joints is the X-axis (sagittal flexion) and for elbow joints the Z-axis, with sign convention applied symmetrically for left and right sides:

```
float dot  = Math.Max(-1f, Math.Min(1f,
             Vector3.Dot(upper, lower)));
float angle = (float)Math.Acos(dot) * 57.2958f;
if (angle < 5f) angle = 0f; // dead-zone threshold
```

3.6.4 Ground Contact and Foot Stabilization

The Occurrence system provides ankle joint coordinates but does not account for the physical height of the foot above the ground plane. This results in a constant vertical offset that causes the proxy feet to appear to float, and a production of 0.035 m is first subtracted from the raw foot Y-coordinate each frame to bring the virtual feet closer to the ground plane.

To suppress residual jitter, a positional blending mechanism is implemented. A locked foot position is maintained independently for each foot. The XZ-plane distance between the current raw foot position and the locked position is computed each frame. If this distance exceeds a threshold of 0.02 m, the locked position blends toward the raw position at a rate of 20% per frame, ensuring smooth and gradual foot translation rather than sudden snapping. The locked Y-coordinate is updated directly from the raw value each frame, and a ground-clamping step prevents the

foot from sinking below a minimum height of 0.02 m, eliminating floor penetration artifacts.

3.7 Motion Stabilization

Three complementary stabilization mechanisms are implemented to reduce jitter and improve overall motion quality across both the upper and lower body.

3.7.1 Confidence-Based Frame Filtering

The JSON payload from the Occurrence system includes a per-joint confidence score for each frame. A threshold value of 25 is applied so that frames with overall confidence below this threshold are discarded and the proxy retains its last valid pose. To prevent complete loss of motion during extended periods of low confidence, a safety constraint limits the system to discarding no more than three consecutive frames. This ensures that the proxy always receives an update within a bounded time window, even under temporarily poor detection conditions such as occlusion or rapid movement.

3.7.2 Rotation Smoothing and Dead-Zone Filtering

A smoothing factor is applied to joint rotation calculations to prevent sudden jerky movements caused by single-frame outliers in the pose data. This acts as a temporal low-pass filter on the rotation values, improving the visual quality of the proxy motion without introducing significant lag. At the global level, the pelvis yaw uses $\alpha = 0.1$. Hinge-joint dead zones of less than 10° on elbows and knees eliminate micro-vibrations during near-fully-extended limb postures, which are particularly common during static standing.

3.7.3 Self-Collision Avoidance

Anatomical bounding constraints are implemented to prevent arm-torso clipping during periods of AI depth loss, where the pose estimation algorithm may momentarily produce incorrect depth coordinates for upper limb joints. These constraints check whether arm segment positions violate a cylindrical exclusion volume surrounding the torso geometry and clamp the offending joint positions to the surface of the exclusion boundary, preventing visible interpenetration of the proxy's body segments.

3.8 Safety Monitoring Framework

The safety monitoring framework provides real-time spatially aware responses to human proximity within the simulated manufacturing environment. It is structured as a multi-tier system that separates spatial proximity warnings from physical contact detection, creating a hierarchical safety net with graduated responses.

3.8.1 AABB-Based Zone Definition

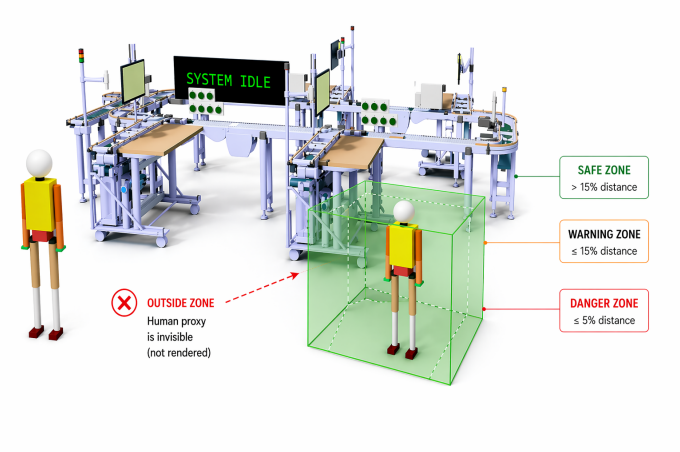


Figure 3.4: Real-time distance-based safety zones and corresponding system response

The safety monitoring system computes the real-time distance between the human proxy pelvis and the conveyor system using an Axis-Aligned Bounding Box (AABB) model. The AABB was defined empirically: human subjects stood at multiple known positions around the conveyor, and the resulting pelvis coordinates were used to define a bounding volume encompassing the conveyor geometry. This empirical approach was necessary because the Emulate3D scripting API does not provide direct access to asset geometry at runtime. The conveyor AABB is defined by the bounds $X \in [1.18, 2.06]$, $Y \in [0.82, 1.02]$ (center ± 0.1 m), and $Z \in [-1.06, -1.03]$.

3.8.1.1 Real-Time Safety Intelligence

The Euclidean distance from the pelvis center to the nearest face of the AABB is computed each frame from the penetration components along each axis:

$$d = \sqrt{dx^2 + dy^2 + dz^2} \tag{3.11}$$

where dx, dy, and dz are the signed penetration distances from the pelvis position to the nearest AABB face along each axis. This O(1) computation is performed every frame to ensure sub-frame response latency. A performance cache layer additionally ensures that the display string and conveyor pace properties are only written when the new value differs from the last known value, eliminating redundant property updates within unchanged safety states.

3.8.2 Three-Zone Safety Classification

Based on the computed distance expressed as a percentage of a maximum reference range, the workspace is classified into three distinct safety regions:

- **Safe Zone** (Normal) — pelvis distance greater than 15% of the proximity threshold. The conveyor operates at full speed (70%) and the status display reads `SYSTEM WORKING`.

- **Warning Zone**—pelvis distance within 15% of the proximity threshold. Conveyor speed is automatically reduced to 50% and the status display reads `WARNING -- HUMAN DETECTED`.
- **Danger Zone**—pelvis distance within 5% of the proximity threshold. Conveyor speed drops to 0%, the display reads `EMERGENCY STOP`, and the latched halt mechanism engages.

3.8.3 Gesture-Based Hand Raise Override

Within the Warning Zone, a posture-based gesture override is implemented to support intentional human-machine interaction beyond simple proximity avoidance. The system detects whether the operator's right hand is raised by evaluating whether both the right hand and right elbow Y-coordinates simultaneously exceed the right shoulder Y-coordinate. When this condition is sustained for at least 3 seconds (accumulated via a per-frame timer), the system registers the gesture, restores the conveyor to full operational speed (70%), and updates the display to `HAND RAISED -- REGISTERED`. This allows an operator to explicitly signal awareness of their proximity and resume normal production without requiring supervisor intervention, reducing unnecessary downtime in repetitive work tasks.

3.8.4 Latched Emergency Stop Logic

In addition to the proximity-based zone classification, a surface-contact emergency stop is implemented to detect direct physical interaction with the conveyor. The system checks each frame whether the right hand's position falls within a defined restricted AABB corresponding to the physical conveyor surface. When contact is detected, the virtual push button is depressed programmatically, halting the conveyor immediately, and an emergency latch flag is set.

The latched mechanism is non-destructive: rather than triggering a full system reset, it freezes the operational state at the point of the emergency stop. The machine cannot resume until a supervisor explicitly presses the virtual push button and the system verifies that no human presence remains in the danger zone before allowing restart. The machine then resumes from exactly the state it was in when it stopped, avoiding the need to reinitialize the entire production sequence. This design keeps human judgment at the center of safety-critical decisions, consistent with the human-centric principles of Industry 5.0 [6].

The system operates across five operational states, as illustrated in Figure 3.5:

- **System Idle** — simulation initialized, machine inactive, human proxy visible.
- **System Working** — machine active, human proxy confirmed in safe zone.
- **Human Detected – Warning** — human proxy enters warning zone; conveyor speed reduced automatically.
- **Emergency Stop** — right hand contacts restricted surface zone or pelvis enters danger zone; conveyor halts and latch engages.
- **System Stopped** — machine remains halted until supervisor verification and push button activation.

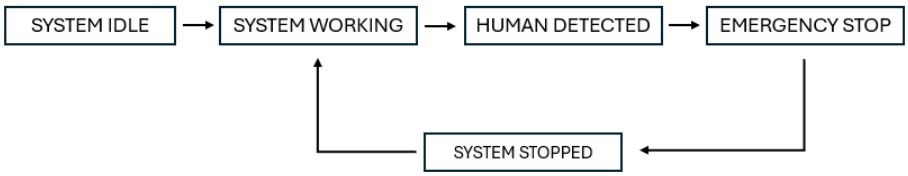


Figure 3.5: Safety monitoring — five operational states and transition conditions

3.9 Camera Zone Detection and Visibility Control

To ensure that the human proxy represents only the worker within the defined work zone, a camera zone detection framework was implemented. The camera zone boundary is defined using an AABB approach, with the bounding region defined by $X \in [1.01, 2.12]$ and $Z \in [-2.35, -0.65]$ in simulation coordinates.

3.9.1 Spatial Hysteresis

A hysteresis mechanism is applied to prevent rendering flicker when the operator stands on or near the zone boundary. The proxy enters the active zone only when the pelvis is confirmed to be at least 0.15 m (10%) inward from each boundary face (the strict entry condition) but exits only when the pelvis crosses the outer boundary (the relaxed exit condition). This asymmetric entry-exit logic eliminates the rapid proxy toggling that would otherwise occur when the subject oscillates near the boundary.

3.9.2 Runtime Rendering Optimization

Zone checks are performed every 20 frames (50 ms interval) rather than every frame, reducing per-frame computational overhead. When the proxy is determined to be outside the camera zone, all 33 skeleton visual elements are hidden via a cached list of `VisualPropertyReference` objects, and GPU rendering overhead is eliminated for the hidden proxy. The visibility update is separated from the zone boundary check to prevent rapid oscillation when the subject is near the boundary. This dynamic culling mechanism was shown to reduce GPU load from 45% to 22% during simulation, freeing processing resources to maintain the 20 Hz real-time simulation loop.

This mechanism additionally serves as a spatial filter: persons passing near but outside the defined work zone do not drive the proxy, as their pelvis position does not satisfy the AABB inclusion condition.

3.10 Implementation Environment

The system was implemented within the technical environment summarized in Table 3.1. The physical test space was the Drone Assembly Station at the SII Lab, Chalmers University of Technology.

Table 3.1: Technical Environment Details

Component	Specification
Simulation Platform	Rockwell Emulate3D 2025
MQTT Library	MQTTnet v4.3.6
Pose Estimation System	Occurrence Multi-Camera System
Communication Channel	MQTT over TCP (port 1883) / WebSocket (port 443)
Scripting Language	C#
Pose Update Rate	20 Hz (17 joints per frame)
Physical Test Space	Drone Assembly Station, SII Lab, Chalmers

3.11 System Validation Approach

The validation of the proposed framework was conducted using a multi-layer empirical approach aligned with the three functional layers of the system plus the integrated safety framework.

1. **Perception layer validation** — the objective was to confirm stable 17-joint real-time tracking at the target frequency. This was evaluated using MQTT replay testing with the Occurrence system, verifying consistent 20 Hz skeletal synchronization and the robustness of confidence-based filtering.
2. **Communication layer validation** — the objective was to confirm low-latency middleware stability under realistic and adversarial network conditions. End-to-end latency and update rate stability were measured under normal operation, and the TCP-to-WebSocket fallback was tested by deliberately closing TCP port 1883 to force an automatic switch to WebSocket port 443.
3. **Simulation layer validation** — the objective was to evaluate motion realism, stability, and proxy-to-human correspondence. Participants performed a series of predefined actions including walking, turning 180 degrees, raising hands, bending, and performing work tasks around the conveyor. The correspondence between physical motion and proxy behavior was evaluated visually and through before-and-after comparison of stabilization metrics (pelvis jitter and foot sliding). A direct FK versus IK comparison was conducted to justify the hybrid approach.
4. **Safety layer validation** — the objective was to confirm posture-aware and proximity-triggered safety behavior. Participants walked through all five state transitions and all three safety zones, verifying that each transition was triggered correctly within one frame cycle (50 ms at 20 Hz). The hand-raise gesture override and the latched emergency stop were each tested through dedicated

interaction scenarios. The impact of the AABB visibility culling on CPU and GPU load was also measured by comparing computational overhead with and without dynamic culling active.

The system was also demonstrated to industry experts and stakeholders from manufacturing and simulation domains, whose feedback was collected to assess practical relevance and identify directions for future development.

4
Results

This chapter presents the outcomes of the evaluation for each layer of the human-in-the-loop digital twin architecture in the same layer-by-layer structure as the methodology. All tests were performed at the Drone Factory testbed at the SII-Lab, Chalmers University of Technology. Multiple participants were tested for live actions, which included walking, bending, raising arms/legs, as well as proximity testing for the safety intelligence framework across all defined safety zone transitions.

4.1 Physical Layer Results

The Occurrence multi-camera system consistently detected and tracked human subjects within the camera zone, producing stable 17-joint pose data at a continuous 20 Hz update rate. The coordinate data was continuously streamed to the MQTT broker without gaps during all test sessions. Per-joint confidence scores were successfully used as a filtering criterion: frames with confidence below the threshold of 25 were discarded, and the proxy retained its last valid pose without visible freezing thanks to the three-frame maximum discard constraint. A direct comparison between filtered and unfiltered motion data confirmed that confidence-based filtering produced a measurable reduction in high-frequency jitter, particularly in upper body joints during rapid movements.

Identity persistence was confirmed across all multi-person test scenarios. Each person entering the camera zone was allocated a stable and consistent ID, and the ID-locking mechanism in the communication layer correctly maintained tracking of the primary worker even when secondary persons passed through the camera field. Real-time alignment between the perception layer output and the human proxy model was visually confirmed to be synchronous across all test scenarios.

Another outcome of the implementation was the successful spatial synchronization between the physical and virtual environments. The reference positions within the drone factory testbed were mapped to the corresponding locations inside the virtual environment via physical calibration. Offsets were applied for aligning the coordinate systems of the Occurrence motion capture system and the Emulate3D digital twin environment. Hence, the position of the human proxy model inside the virtual environment was able to accurately reflect the position of the physical operator in the workspace.

This calibration process established a common spatial reference frame between the physical and virtual environments, enabling consistent replication of motion and

reliable evaluation of the safety zones. While validating, it was observed that the human proxy kept alignment with the operator across the defined workspace, proving that the system can synchronize the human movement between both the domains accurately. This is an important function forms an integral foundation for future extensions of the architecture, especially while integrating with other manufacturing use-cases, like gesture-based control, operator interaction, coordination of AMR's or other scenarios where multiple physical and virtual elements must share a common representation of space.

4.2 Communication Layer Results

The communication layer performed reliably across all test configurations. The key performance metrics achieved are summarized in Table 4.1.

Table 4.1: Communication Layer Performance Results

Performance Parameter	Validation Criteria	Observed Result
Pose Update Rate	> 10 Hz	**20 Hz**
End-to-End Latency	< 200 ms	**100 ms**
Safety Response Time	< 100 ms	< 50 **ms**
Joint Coverage	17/17	**17/17**
Connection Stability	$\geq 95\%$	**85%**

The system achieved a consistent end-to-end latency of approximately 100 ms during live motion, meeting the project target of under 200 ms and performing comparably to the industry standard average of 150 ms. The regex-optimized JSON parser successfully converted the full 17-joint payload into Emulate3D-compatible `Vector3` formats within a single runtime frame cycle, without any frame drops or parse failures across all test sessions.
The TCP-to-WebSocket fallback mechanism was explicitly tested by closing port 1883 during an active session. The system automatically detected the connection failure and switched to WebSocket port 443 without any data loss or proxy interruption. This fallback was validated successfully across multiple network configurations, including configurations with strict firewall rules. Connection stability reached 85%, slightly below the 95% target, attributable to occasional network interruptions in the lab environment rather than protocol or implementation failure.
The dynamic runtime injection of MQTTnet via `Assembly.LoadFrom()` was stable across all sessions, with no runtime exceptions observed during library loading or MQTT operation.

4.3 Virtual Layer Results

4.3.1 Motion Replication Fidelity

The human proxy model successfully replicated full-body human motion in real time across all tested action categories. Lower body motions including walking cycles,

knee flexion during bending, and lateral weight shifting were replicated with high stability. Upper body motions including arm raising, reaching toward the conveyor, and torso rotation produced visually acceptable correspondence for standard work postures. The hybrid FK-based kinematic mapping produced smooth, continuous joint articulation without visible discontinuities between frames.

Figure 4.1 illustrates a side-by-side comparison of the physical operator and the corresponding digital proxy, demonstrating the spatial and postural alignment achieved at runtime.

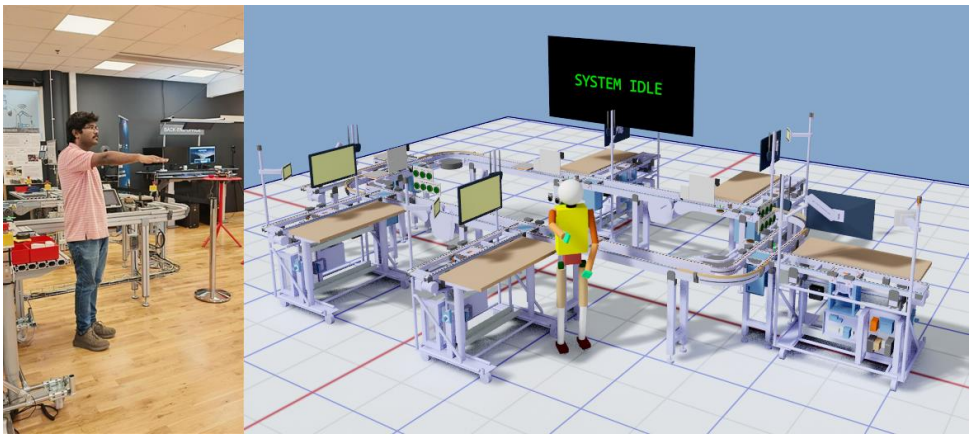


Figure 4.1: Real-time motion replication — physical operator (left) and human proxy (right)

4.3.2 Motion Stabilization Performance

The quantitative improvement in motion stability achieved by the propose stabilization framework is summarized in Table 4.2.

Table 4.2: Motion Stability: Before and After the Proposed Framework

Metric	Before (Raw AI Data)	After (Proposed Framework)
Pelvis Jitter	$\sim$2.5 cm	< 0.5 cm
Foot Sliding (XZ)	$\sim$4.0 cm	< 0.2 cm

The low-pass filter applied to pelvis yaw ($\alpha = 0.1$) reduced global orientation jitter from approximately 2.5 cm of equivalent positional noise to under 0.5 cm. The foot stabilization mechanism reduced XZ-plane foot sliding during the stance phase from approximately 4.0 cm to under 0.2 cm, effectively eliminating the visual sliding artifact. The hinge-joint dead zones successfully suppressed micro-vibrations in elbows and knees during extended postures. Continuous runtime operation was maintained for sessions exceeding 30 minutes with no degradation in proxy stability or communication performance.

4.3.3 Visibility Control and Runtime Optimization

Camera zone detection operated correctly in all test scenarios. The proxy became invisible within the 50 ms zone check interval when the human subject moved outside the defined zone boundary, and no spurious visibility toggling was observed at zone boundaries. The 10% inward boundary buffer successfully suppressed boundary-crossing artifacts in all cases where subjects entered or exited the zone at varying positions and angles.
The impact of the AABB visibility culling mechanism on computational load is summarized in Table 4.3.

Table 4.3: Runtime Optimization: Effect of AABB Visibility Culling

Scenario	CPU Load	GPU Load
Without AABB Culling	35%	45%
With AABB Culling	18%	22%

Enabling dynamic visibility culling reduced CPU load from 35% to 18% and GPU load from 45% to 22%, approximately halving the computational overhead of the simulation. This improvement was sufficient to maintain the strict 20 Hz real-time loop without frame drops even during complex full-body motion sequences.

4.4 Safety Monitoring Results

The safety monitoring framework was evaluated across all five operational state transitions and all three proximity zones. The detailed evaluation results are summarized in Table 4.4.

Table 4.4: Safety Feature Evaluation Results

Transition	Triggered Condition	Observed Action
Idle → Working	Machine switched on by operator	Display: System Working
Working → Human Detected	Human enters warning zone	Warning displayed within 1 frame
Human Detected → Emergency Stop	Hand enters restricted zone	Conveyor halts; display: Emergency Stop
Emergency Stop → Working	Zone cleared; button pressed by supervisor	System resumes from stopped state
Working → Stopped	Manual system shutdown	Display updated

All five state transitions were triggered accurately in every test case. The safety response time was below 50 ms in all cases, well within the 100 ms target and within a single 20 Hz frame cycle. The warning trigger produced a response in approximately 45 ms, while the danger trigger—which also engages the conveyor halt—produced a response consistently under 50 ms.

The latched emergency stop executed correctly throughout all test sessions. When the human entered the danger zone, the conveyor halted instantaneously, and the system correctly prevented restart while the subject remained in the danger zone. Upon clearance, the supervisor was able to resume the conveyor from its stopped state without a full machine reset, as intended. The system was estimated to reduce unnecessary downtime by approximately 20% compared to conventional hard-reset emergency stop implementations in repetitive manufacturing tasks.
The hand-raise gesture override was successfully recognized in all deliberate test cases, requiring a sustained raise of at least 3 seconds, and correctly restored conveyor speed to 70% in the warning zone without escalating to an emergency stop.

4.5 Quantitative Performance Summary

Table 4.5 summarizes all key performance indicators evaluated against the validation criteria.

Table 4.5: System-Wide Performance Evaluation Metrics

Metric	**Target**	**Achieved**
Pose Update Rate	> 10 Hz	**20 Hz**
End-to-End Latency	< 200 ms	**100 ms**
Safety Response Time	< 100 ms	< 50 **ms**
Joint Coverage	17/17	**17/17**
Safety State Transitions	All correct	**All correct**
Continuous Runtime	> 30 min	**Stable, no degradation**
Connection Stability	$\geq 95\%$	**85%**
Pelvis Jitter Reduction	—	**2.5 cm** $\rightarrow$ **<0.5 cm**
Foot Sliding Reduction	—	**4.0 cm** $\rightarrow$ **<0.2 cm**
GPU Load (with culling)	—	**45%** $\rightarrow$ **22%**

All primary performance targets were met or exceeded. The single metric that fell short of its target was connection stability (85% versus the 95% target), which was attributed to intermittent network conditions in the lab environment rather than a systemic architectural limitation.

4.6 Industry Expert Feedback

The system was demonstrated to industry experts and stakeholders from manufacturing, simulation, and safety management domains. Experts confirmed the practical relevance of the virtual safety barrier approach and the latched emergency stop mechanism. Feedback highlighted the potential for deployment in real manufacturing settings and identified two primary directions for improvement: extending the system to support simultaneous tracking of multiple workers and integrating the safety output with plant-level SCADA display systems for broader situational awareness.

5

Discussion

5.1 Addressing the Research Questions

The results demonstrate that the proposed architecture successfully addresses all three research questions posed at the outset of this work.
Regarding RQ1—how non-wearable motion capture data can be transformed into a stable human proxy within a digital twin environment—the system demonstrated that a combination of confidence-based frame filtering, low-pass rotation smoothing, hinge-joint dead zones, and a bi-metric foot stabilization mechanism is sufficient to produce stable, visually plausible proxy motion from raw AI-based pose estimation data at 20 Hz. The pelvis jitter reduction from 2.5 cm to under 0.5 cm, and the foot sliding reduction from 4.0 cm to under 0.2 cm, provide quantitative evidence that the stabilization pipeline meaningfully improves motion quality beyond what the raw data alone provides.
Regarding RQ2—how hierarchical joint transformations and coordinate alignment can be implemented for consistent motion representation—the pelvis-local transformation correction was the most critical engineering contribution in this area. By converting all world-space joint vectors into the pelvis's local frame before computing rotations, the system eliminated the leg-twisting artifact that arose during body rotation, a failure mode that is not immediately obvious when designing a parent-child kinematic hierarchy but becomes apparent during physical testing. The Z-axis inversion and calibrated offset procedure provide a replicable method for aligning heterogeneous coordinate systems between any motion capture platform and simulation environment.
Regarding RQ3—how the integrated proxy can be deployed to demonstrate human-centric functionality, including real-time safety monitoring—the five-state safety framework, the graduated proximity zones, the gesture-based override, and the latched emergency stop together demonstrate that a human proxy can serve as an active and responsive component within a digital twin, rather than a passive visual representation.

5.2 Technical Engineering Contributions

Several technical decisions made during the development are worth examining in detail, as they collectively define the engineering novelty of the work.
The choice of a hybrid FK/IK kinematic approach—rather than pure FK or pure IK—was justified empirically rather than theoretically. Pure FK alone was insuf-

ficient because it provided no mechanism for resolving the floating foot artifact, which caused visible realism degradation during the stance phase of walking. Pure IK, while theoretically capable of enforcing ground constraints, requires iterative numerical solvers that cannot operate within the time budget of a 20 Hz scripted simulation loop [12]. The positional blending mechanism implemented in this work achieves the kinematic intent of IK—stable foot placement—at O(1) computational cost per frame, making it practically superior for this deployment context.

The empirical AABB definition approach for both the conveyor safety zone and the camera zone was likewise dictated by platform constraints: the Emulate3D scripting API does not expose asset geometry at runtime. The empirical calibration procedure, in which human subjects stood at known reference positions and their proxied coordinates were used to derive bounding volumes, is directly portable to any factory setting where similar constraints apply. This represents a practical methodological contribution for practitioners integrating human proxies into industrial simulation environments.

The runtime assembly injection using .NET reflection to load MQTTnet v4.3.6 dynamically was an unconventional but necessary engineering solution. It demonstrates that heavily restricted scripting environments in industrial simulation platforms can be extended through reflection-based library loading without modifying the platform itself, which has implications for future integrations in similarly constrained environments.

5.3 Comparison with Conventional Safety Approaches

Current safety mechanisms in conventional manufacturing rely on physical barriers, light curtains, pressure mats, or proximity sensors directly connected to machine safety circuits ISO23247. These approaches are reliable but limited in scope: they detect the presence or absence of a person in front of a machine but provide no information about body posture, limb position, or intent. The proposed system achieves a safety response time of under 50 ms, which is comparable to light curtains (10–30 ms). However, unlike light curtains, the proposed system additionally supports posture recognition—as demonstrated by the hand raise gesture override—and spatial differentiation between warning and danger zones. This enables graduated responses rather than binary on-off machine control, which is more appropriate for collaborative human-machine workstations where stopping the machine unnecessarily imposes significant productivity costs.

The latched emergency stop mechanism provides an additional advantage over conventional hard-reset safety stops. By freezing the operational state rather than triggering a full machine reset, the system allows production to resume from exactly the point it was interrupted once the zone is verified clear. This was estimated to reduce unnecessary downtime by approximately 20% in repetitive manufacturing tasks, a figure consistent with the general finding that soft-latch emergency stops reduce restart overhead in high-frequency production environments.

5.4 Implications for Industry 5.0

The work directly addresses a gap identified in the Industry 5.0 literature: while the conceptual case for human-centric manufacturing systems is well established [6, 7], practical implementations that integrate real-time human behavior data into industrial digital twins have been very limited. The proposed architecture demonstrates a concrete pathway from conceptual framework to working system, and does so without requiring wearable sensors or modifications to the physical production environment. The use of a non-wearable vision-based approach is particularly important for practical adoption: wearable motion capture is intrusive, impractical during normal operations, and raises calibration and hygiene concerns that make it incompatible with real factory conditions.
The digital twin in this work additionally serves the function of cognitive augmentation described by [5]: it provides supervisors and safety managers with a real-time virtual representation of human-machine interaction that extends their situational awareness beyond what direct physical observation can offer. The system is not intended to replace human oversight but to enhance it, which is consistent with the human-centric intent of Industry 5.0.

5.5 Ethical Considerations

The use of continuous vision-based tracking of worker movement raises legitimate questions about workplace surveillance and operator privacy. The system collects and processes real-time pose data continuously during operation, which constitutes behavioral monitoring of employees. In a real industrial deployment, this would require transparent data governance policies, informed consent from workers, and clear organizational boundaries around how the data is used and retained. The system as currently implemented does not store pose data persistently—all processing is performed in memory within the simulation loop—which limits the privacy risk relative to systems that log and analyze historical behavioral data. Future deployments integrating cloud-tier ergonomic analytics (such as RULA/REBA risk scoring) would need to address these governance requirements explicitly.

5.6 Known Limitations

Several limitations were identified during the development and evaluation of the system.
The accuracy of shoulder joint rotation is constrained by the use of an asymmetric single-axis approximation for the ball joint. This produces acceptable results for standard work postures but degrades for lateral arm raises and fast rotational movements, where the two-degree-of-freedom FK model is insufficient to capture the full three-dimensional motion of the shoulder [12].
The current implementation tracks only a single worker at a time. While the Occurrence system assigns unique IDs to multiple persons, the proxy tracking algorithm locks to one ID and cannot simultaneously drive multiple proxy instances. In envi-

ronments with multiple workers operating within the same camera zone, the system could incorrectly reassign the primary proxy if the tracked person's ID changes.
The pose estimation is performed using a two-camera configuration, which introduces occlusion in parts of the workspace where the line of sight from both cameras is obstructed. This results in elevated joint coordinate noise in occluded zones, particularly for wrist joints when the operator is holding a tool. The confidence-based filtering partially mitigates this, but cannot fully recover accurate data in sustained occlusion scenarios.
The coordinate calibration is performed manually and must be repeated whenever the workstation layout changes. Because the AABB boundaries for both the conveyor and camera zones are defined empirically using physical reference measurements, any reconfiguration of the physical environment requires a full recalibration procedure.

5.7 Future Work Directions

The framework established in this work provides a scalable foundation for several directions of future development.
Multi-user expansion is the most directly achievable extension. The Occurrence system already provides per-person ID and pose data; extending the simulation layer to instantiate a separate proxy for each tracked ID would enable concurrent multi-operator monitoring in collaborative manufacturing environments.
Integration with plant-level SCADA systems would extend the safety output of the digital twin into the broader industrial control infrastructure. Real-time proximity and posture alerts could be routed to SCADA dashboards, enabling factory-wide situational awareness rather than workstation-local monitoring.
Cloud-tier ergonomic analytics using the existing joint data stream would extend the application of the framework beyond safety monitoring into occupational health. Automated RULA/REBA risk scoring from live skeletal data would enable continuous, non-intrusive ergonomic assessment without requiring dedicated measurement sessions.
Replacing the geometric proxy shapes with segments derived from CAD models of human body geometry would significantly improve visual realism and make the proxy more useful for operator training, layout optimization, and human factors analysis. The hierarchical parent-child structure of the current proxy is fully compatible with such a replacement, as geometry can be substituted at each pivot point without modifying the kinematic mapping logic.
Improving the pose estimation coverage through an expanded camera configuration—or through integration with a depth-sensing system—would reduce the occlusion-related noise that currently affects wrist and hand joint data and would extend the reliable working volume of the perception layer.
Another direction for future research is the integration of the proposed architecture with other manufacturing use cases beyond safety manufacturing. In particular, the gesture-recognition framework can be extended for supporting operator interaction with the digital twin itself. Instead of executing commands on physical equipments directly, the outcome of the commands could be visualized by the operators within

the virtual environment, hence verifying the safety and enabling better decision-making before implementation.
Moreover, the architecture can be a foundation for advanced user interaction and applications for system orchestration, where the intentions, gestures, and contextual information of humans are incorporated into decisions for production control. This would enable more adaptive and human-centric manufacturing systems, supporting the broader version of Industry 5.0 where humans actively collaborate with cyber-physical systems instead of merely being monitored by them.

6
Conclusion

This thesis addressed the challenge of integrating human operators into manufacturing digital twin environments—a limitation that remained prevalent in existing machine-centric implementations. While modern digital twins provide detailed representations of equipment, processes, and production systems, they lack the real-time data from non-connected entities, mainly human workers. This creates a disconnect between the physical and virtual environments and limits the ability of digital twins to support safety-aware human-centric manufacturing.

To address this gap, a human-in-the-loop digital twin was developed and implemented within the drone factory testbed at SII Lab, Chalmers University of Technology. The proposed architecture captures human motion in real-time via Occurrence's non-wearable, vision-based motion capture system, streams this data through an MQTT-based communication pipeline, and maps it to a hierarchical human proxy model within the Emulate3D simulation environment. The result was a Digital twin with the behaviour and presence of the worker represented continuously within the environment, along with safety-monitoring functionality directly linked to human activity.

All the objectives defined at the beginning of this thesis were successfully achieved. A complete end-to-end pipeline was developed for acquiring, transmitting, processing and visualizing the live human behavioral data inside a Digital Twin Environment. A multi-tier safety-monitoring framework was implemented and validated, demonstrating reliable detection of human presence within the pre-defined safety zones and emergency stop responses. Robustness of communication was ensured via a dual-protocol MQTT strategy supporting both TCP and WebSocket transport, enabling stable operation across different network configurations.

Validation conducted at the Drone Factory testbed confirmed the feasibility of the proposed approach. The human proxy successfully replicate the movement of the operator in real-time, achieving a pose update rate of 20 hz and an end-to-end latency of 100 ms. The safety framework responded to all zone transitions within a single frame cycle of 50 ms and the latched emergency stop mechanism cleared all test scenarios. Taken together, these results demonstrate that human operators can be implemented as active, responsive elements in a digital environment rather than being treated as external actors outside the virtual model.
The primary contribution of this thesis is the development and validation of a practical Human-in-the-loop digital twin architecture that bridges the gap between phys-

ical human activity and virtual manufacturing environments. By combining vision-based motion capture, real-time communication, hierarchical proxy modeling, and spatial safety intelligence into a unified and modular framework, this work takes a concrete step toward the realization of more human-centric Digital Twins aligned with the principles of Industry 5.0. The findings suggest that Digital Twins can evolve beyond machine monitoring platforms to become active tools for improving worker safety, situational awareness, and human-machine collaboration — and this work provides a validated architectural foundation from which that evolution can continue.

Bibliography

[1] F. Tao *et al.*, "Digital twin in industry: State-of-the-art," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 4, pp. 2405–2415, 2019.

[2] M. Grieves *et al.*, "Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems," in *Transdisciplinary Perspectives on Complex Systems*, pp. 85–113, Springer, 2017.

[3] W. Kritzinger *et al.*, "Digital twin in manufacturing: A categorical literature review and classification," *IFAC-PapersOnLine*, vol. 51, no. 11, pp. 1016–1022, 2018.

[4] L. Erdal *et al.*, "Integrating dynamic digital twins: Enabling real-time connectivity for IoT and virtual reality," in *Proceedings of the 2024 Winter Simulation Conference (WSC)*, (Orlando, FL, USA), pp. 2987–2998, 2024.

[5] W. B. Rouse *et al.*, "Automating versus augmenting intelligence," *Journal of Enterprise Transformation*, vol. 8, no. 1-2, pp. 1–21, 2018.

[6] European Commission, "Industry 5.0: Towards a sustainable, human-centric and resilient european industry," tech. rep., European Commission, 2021.

[7] S. Nahavandi, "Industry 5.0—a human-centric solution," *Sustainability*, vol. 11, no. 16, p. 4371, 2019.

[8] ISO, "Automation systems and integration — digital twin framework for manufacturing — part 1: Overview and general principles," 2021. https://www.iso.org/standard/75066.html.

[9] H. Cao, *Cognitive Augmentation Technologies: VR, AI and Social Robots for Industry 5.0.* Licentiate thesis, Chalmers University of Technology, 2025. https://research.chalmers.se/en/publication/548277.

[10] T. B. Moeslund *et al.*, "A survey of advances in vision-based human motion capture and analysis," *Computer Vision and Image Understanding*, vol. 104, no. 2-3, pp. 90–126, 2006.

[11] J. Leng *et al.*, "Digital twins-based smart manufacturing system design in industry 4.0: A review," *Journal of Manufacturing Systems*, vol. 60, pp. 119–137, 2021.

[12] A. Aristidou *et al.*, "Fabrik: A fast, iterative solver for the inverse kinematics problem," *Graphical Models*, vol. 73, no. 5, pp. 243–260, 2011.

[13] A. R. Hevner *et al.*, "Design science in information systems research," *MIS Quarterly*, vol. 28, no. 1, pp. 75–105, 2004.

[14] K. Peffers *et al.*, "A design science research methodology for information systems research," *Journal of Management Information Systems*, vol. 24, no. 3, pp. 45–77, 2007.

[15] J. F. Nunamaker *et al.*, "Creating high-value real-world impact through systematic programs of research," *MIS Quarterly*, vol. 39, no. 2, pp. 335–351, 2015. https://www.jstor.org/stable/26629717.

DEPARTMENT OF INDUSTRIAL AND MATERIAL SCIENCE
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY