

PAPER • OPEN ACCESS

## Supporting Ergonomics Evaluations in Manufacturing – A Comparison of Computer Vision- and IMU-Based Motion Capture

To cite this article: Raquel Quesada Díaz *et al* 2026 *IOP Conf. Ser.: Mater. Sci. Eng.* **1342** 012053

View the [article online](#) for updates and enhancements.

### You may also like

- [An investigation of low ergonomics risk awareness among staffs at early product development phase in Malaysia automotive industries](#)  
Fazilah Abdul Aziz, Noraini Razali and Nur Najmiyah Jaafar
- [Research on the Application of Computer Ergonomics in Industrial Product Design](#)  
Xin Fan
- [A Benchmarking of the Integrated Train Driver Performance Model](#)  
Jalil Azlis-Sani, Siti Zawiah Md Dawal and Norhayati Mohamad Zakwan

# Supporting Ergonomics Evaluations in Manufacturing – A Comparison of Computer Vision- and IMU-Based Motion Capture

Raquel Quesada Díaz<sup>1</sup>, Aitor Iriondo Pascual<sup>1</sup>, Dan Högberg<sup>1</sup>, Sunith Bandaru<sup>1</sup> and Lars Hanson<sup>1,2</sup>

<sup>1</sup>School of Engineering Science, University of Skövde, Sweden

<sup>2</sup>Department of Engineering, Volvo Construction Equipment, Arvika, Sweden

E-mail: raquel.quesada.diaz@his.se

**Abstract.** Ergonomics evaluation methods are crucial for assessing risks for work-related musculoskeletal disorders and ensuring operator well-being, productivity, and safety. Despite the increased use of digital twins and AI-supported tools in production system design and operation, ergonomics evaluations still primarily rely on observational techniques such as expert assessments or checklist-based tools like the Rapid Entire Body Assessment and the Rapid Upper Limb Assessment. These methods are time-consuming, imprecise, and prone to subjectivity arising from variability in the judgment of the ergonomists and ambiguity in scoring criteria. As an alternative, ergonomics evaluation methods based on using technologies for direct measurements can provide semi-automation of the assessments and offer greater objectivity and precision. This study investigates the capability of a computer vision-based motion capture approach to support direct measurement ergonomics evaluations and compares its results with those of an inertial measurement unit-based system in an industrial task. The comparison was conducted by studying output data of the two systems and by feeding the data into a direct measurement-based ergonomics evaluation method. A representative industrial assembly task involving upper-body movement and dynamic wrist activity was recorded simultaneously using a single monocular RGB camera and an IMU-based system. Both datasets were processed using two parallel workflows that followed the same structure to extract joint angles and segment positions over time. The comparison and evaluation of the results demonstrates that computer vision-based motion capture has the potential to provide human posture and motion data suitable for direct measurement ergonomics evaluations in industrial environments.

## 1 Introduction

Industry 4.0 [1] represents the digital transformation of manufacturing, driven by the integration of cyberphysical production systems, the Industrial Internet of Things, and advanced data analytics, among other technologies [2]. These technologies enable machines, devices, and information systems to communicate and cooperate in real time, creating connected, adaptive, and self-optimising production environments [3, 4]. This combination of digital and physical processes has improved productivity, flexibility and quality across manufacturing networks [5, 6]. However, the emphasis on automation and efficiency has also raised concerns about workforce



displacement, new skills demands, and limited attention given to human well-being in highly automated environments [7]. This imbalance led the European Commission to introduce the Industry 5.0 [8] vision, which builds on existing people-centred approaches such as Lean production by aligning technological progress with human well-being, resilience, and sustainability, so that value is created for both companies and society [8, 7].

As this vision takes shape, the manufacturing industry is transitioning towards a sustainable and human-centred one where the role of the human worker is being redefined. The concept of Operator 5.0 [9], describes the future operator as an empowered professional equipped with digital competencies and intelligent tools that enable seamless interaction with advanced technologies. Operators are no longer viewed as passive assets, but rather as active contributors within intelligent, adaptive, collaborative, and resilient production systems. This shift demands digital tools that model, simulate, analyse and optimise production environments while prioritising operator well-being, task inclusivity and long-term employability. In this context, accurate and efficient methods for assessing physical workload and musculoskeletal risk are fundamental to designing and operating production systems that are both productive and sustainable.

Ergonomics evaluation methods are essential for assessing the risks for work-related musculoskeletal disorders (WMSDs) and for ensuring operator well-being, productivity and safety. However, despite the increased use of digital technologies and AI-supported tools in production system design and operation, ergonomics evaluations still primarily rely on observational techniques, such as expert assessments or checklist-based tools like the Rapid Entire Body Assessment (REBA) [10] and the Rapid Upper Limb Assessment (RULA) [11]. Albeit useful and commonly used, research has shown that observation-based ergonomics evaluation methods are time-consuming, imprecise and susceptible to subjectivity due to variability in the judgment of ergonomists and ambiguity in scoring criteria [12, 13]. Recent advances in motion capture (MoCap) technologies have enabled new developments in ergonomics research, especially in direct measurement and digital risk assessment. Among the digitalised ergonomics evaluation methods, the Lund Action Levels (LAL) [14] represents a data-driven approach for assessing musculoskeletal risk of joint angles and angular velocities. The method provides percentile-based action levels derived from field measurements, enabling the integration with MoCap-based evaluation workflows.

## 2 Related Work

Direct measurement methods refer to the use of technology to collect physical exposure data directly from the operator's body through either sensor- or image/vision-based technologies. Sensor-based approaches include digital goniometers [15, 16], Inertial Measurement Units (IMUs) [17, 18], Electromyography (EMG) sensors [15, 19] or depth-sensing cameras such as Microsoft Kinect [20, 16], which record kinematic, muscular or physiological data via contact or wearable devices. In contrast, image- or vision-based approaches, rely on optical data acquisition and computational analysis, ranging from photogrammetry [18, 21] deep learning-based computer vision (CV) systems [22, 23], to estimate postures and motions from video or image data. Both approaches aim to reduce subjectivity and increase objectivity, precision, and the potential for automation; however, they differ in accuracy, costs, expertise needed, intrusiveness level, and usability, highlighting the need for more accessible and practical solutions [21, 23].

To address these challenges, four main MoCap technologies have been explored in ergonomics research:

1. **Optoelectronic systems**, which use advanced cameras, infrared and reflective markers to reconstruct precise three-dimensional (3D) movements [24] (e.g. Qualisys, Move4D, Vicon).
2. **Deep-sensor systems**, which derive 3D skeletons from structured-light or time-of-flight depth sensing [19] (e.g. Microsoft Kinect, Intel RealSense).
3. **CV-based systems**, which estimate human posture and motion from RGB images using AI-based models [25, 26] (e.g. BlazePose, OpenPose, DeepPose, AlphaPose).
4. **IMUs-based systems**, which employ wearable inertial sensors to calculate segment orientations and joint angles [17, 27] (e.g. Xsens, Wergonics).

Several studies have compared different MoCap technologies in terms of accuracy, repeatability, and suitability for human motion (gait) analysis. Early validation work done by Ceseracciu et al. [24] compared Markerless camera-based systems (MCBS) and optical Marker-based systems (MBS) through simultaneous gait data collection. Their results showed that sagittal plane kinematics could be estimated with reasonable accuracy. However, the estimation for the transverse and frontal planes were less precise due to calibration and viewpoint sensitivity. Das et al. [28] built on this study to propose an statistical framework that could quantify agreement between MCBS and MBS solutions using functional limits of agreement (fLoA). The study demonstrated that even though MCBS were accurate capturing general patterns, there was a substantial bias between the measurement of joint-angle amplitudes that required post-processing for removal. A comprehensive literature review by Scataglini et al. [23] evaluated 22 studies comparing MCBS (e.g. Kinect, Motognosis Labs, Theia3D) with MBS (e.g. Vicon, Optotrack, Qualisys) for gait analysis. The meta-analysis concluded that 3D MCBS showed good reliability for spatiotemporal parameters such as stride time and walking speed, among others. Sagittal-plane kinematics were the most accurate for hip and knee analysis. Nonetheless, the authors emphasised the need for standardisation, and reported poor concurrence in the transverse and frontal planes highlighting the need for further system evaluation, which are relevant limitations to industrial settings with non-ideal camera angles and occlusions.

Some studies have compared sensor-based and hybrid approaches. Lind et al. [27] for example, demonstrated the potential of IMUs-based “smart workwear” for continuous monitoring of physical exposure in real work environments. Their textile-integrated system aimed to prevent work-related musculoskeletal disorders WMSDs by embedding sensors directly into clothes, illustrating the industrial feasibility of IMUs-based solutions. Ciccarelli et al. [19] extended this approach by proposing an automated framework that integrated IMUs, EMG sensors, and other wearable devices for real-time ergonomic risk assessments in Industry 5.0 contexts. Their findings emphasised the advantages of sensor fusion for accuracy and automation but noted the remaining challenges in calibration and user comfort. Other works have directly compared IMUs- and camera-based systems. Meletani et al. [21] evaluated a 4D stereophotogrammetry system (Move4D) against Xsens MVN Awinda (Xsens) for gait analysis. The authors stated that Move4D showed high agreement with Xsens for the spatiotemporal parameters and moderate differences the kinematic data for the joint angles, i.e. the system showed reliability and accuracy comparable with Xsens. A similar comparison between IMUs-based and 4D stereophotogrammetry (Move4D) was done by Fontinovo et al. [18] for the RULA ergonomics risk assessment method. RULA results were comparable across methods, although there were minimal differences present that hinted the existence of some discrepancies between the systems due to differences in joint-angle capture and measurement definitions. Both

studies [21, 18] demonstrate that optical and inertial systems can produce comparable ergonomic insights, provided that calibration and methodological consistency are ensured.

Recent research has increasingly focused on comparing IMUs- and CV-based motion capture systems to evaluate their suitability for ergonomic risk assessment and posture and motion estimation. Elango et al. [26] compared CV-based systems with wearable IMUs in an automotive production case, combining depth cameras and inertial sensors for posture evaluation. The results indicated that while CV-based systems could reproduce posture trends with fewer sensors, hybrid configurations achieved higher accuracy. In a subsequent study [25] the authors introduced ERAIVA, a video-based ergonomic posture evaluation software, and benchmarked it against conventional manual assessments. The study confirmed the feasibility of CV-based systems to identify awkward postures but noted that accuracy still depended on viewing angle and task dynamics. Similarly, Agostinelli et al. [29] validated low-cost 2D RGB camera-based MoCap tools across multiple manufacturing environments. The findings showed that these are promising for continuous posture monitoring but limited by environmental variability and calibration compared to wearable IMUs. These studies emphasise that while IMUs-based systems remain the reference standard due to their robustness and independence from lighting or occlusion, CV-based systems provide a non-intrusive, cost-efficient alternative suitable for scalable workplace applications.

Table 1 shows a summary of the previous comparative studies on MoCap technologies.

The present study builds on previous work integrating MoCap with digitalised ergonomics evaluations in industrial settings [17], and investigates whether a monocular (single-camera) CV-based MoCap approach can provide valid motion data to support direct measurement ergonomics evaluations and compares its results with those of an IMUs-based system in an industrial task. The comparison of CV- and IMUs-based motion data using this framework enables a rigorous evaluation of the feasibility of camera-based ergonomics assessments in industrial settings.

### 3 Method

The study followed the workflow in Figure 1, comprising three stages: data collection, data processing, and data analysis. The following sections will describe each step in the workflow in further detail.

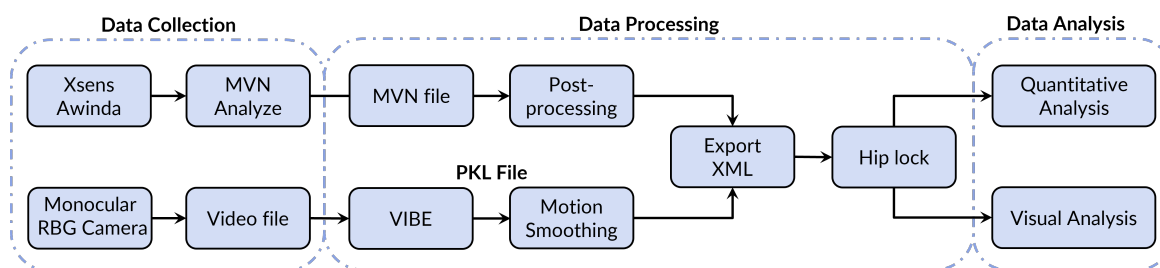


Figure 1: Pipeline for data collection, data processing, and comparative analysis.

#### 3.1 Experimental Setup

The data collection was conducted in a mock-up battery assembly robot cell at ASSAR Industrial Innovation Arena [30], Skövde, Sweden. The cell (Figure 2) features a collaborative robot mounted on a modular metal workstation, providing a flexible environment for assembly-related operations. Components were arranged to emulate a realistic industrial assembly scenario. The

Table 1: Summary of previous comparative studies on motion-capture technologies.

Author (Year)	Study / Context	MoCap Systems Compared	Domain	Main Findings / Relevance
Ceseracciu et al. (2014)	Markerless vs. marker-based optical MoCap during gait	Optical (markerless) vs. Optical (marker-based)	Gait	Sagittal-plane kinematics accurate; poor reliability in transverse/frontal planes.
Das et al. (2023)	Functional limits of agreement (fLoA) framework	Markerless vs. marker-based	Gait	Captured general movement patterns accurately but systematic amplitude biases remained.
Scataglini et al. (2024)	Systematic review and meta-analysis (22 studies)	MCBS vs. MBS	Gait	High temporal accuracy (ICC 0.81–0.98); poor validity for non-sagittal planes; need for standardised protocols.
Elango et al. (2022)	Evaluation of posture in automotive production tasks	Depth camera (CV) vs. IMU	Industrial case	AI-based CV system reproduced posture trends; hybrid IMU–CV setup achieved highest accuracy.
Elango et al. (2024)	Evaluation of ERAIVA video-based risk assessment tool	CV (RGB video) vs. manual/IMU reference	Workplace ergonomics	Demonstrated high usability; accuracy influenced by camera angle and motion complexity.
Agostinelli et al. (2024)	Validation of low-cost CV-based MoCap in manufacturing	2D RGB CV-based vs. literature IMU references	Manufacturing	Promising for continuous posture monitoring; sensitivity to lighting and environment noted.
Lind et al. (2019)	Smart workwear concept for exposure monitoring	IMU-based textile sensors	Workplace	Demonstrated feasibility of continuous ergonomic monitoring through wearable IMUs.
Cicarelli et al. (2025)	Automated sensor-based risk assessment	IMU + EMG + other wearables	Industry 5.0	Framework for automated risk assessment; accuracy depends on calibration and comfort.
Meletani et al. (2024)	Comparison of 4D stereophotogrammetry vs. IMU	Move4D vs. Xsens	Gait	High agreement for timing parameters; minor joint-angle differences.
Fontinovo et al. (2025)	Comparison of ergonomic risk assessments (RULA)	Observational vs. IMU vs. 4D stereophotogrammetry	Ergonomics	RULA score differences due to joint-angle measurement, not risk model; confirms comparability of methods.

setup is part of the ASSAR Research Lab and Technology Labs, which provides an environment for applied research and technology demonstration in close collaboration between academia and industry partners.



Figure 2: Mock-up assembly station consisting of a battery assembly robot cell, a table, and rearrangeable objects.

The workstation included a worktable with tools and components placed within the operator's reach: two triangular metal brackets fitted with screws and nuts, two standard hand tools (a hammer and a wrench), and a plastic container containing assembly parts weighing a total of 10 kg. The triangular brackets could be attached to the robot frame to reconfigure the assembly, enabling realistic manipulation and fastening operations. The setup was designed to reproduce task variability typical of industrial assembly operations, including reaching, grasping, lifting, aligning, manipulating, fastening, hammering, positioning, and handling components of different sizes and weights.

For the purpose of this study, the first author served as the operator at the workstation. The operator was an adult female with experience in industrial assembly and prior training in the use of motion-capture systems. The operator participated voluntarily, and no formal consent procedure was required since the study involved self-experimentation. The experiment complied with institutional and national ethical guidelines for research involving human participants.

### 3.2 Task Design

Two work tasks were defined to reproduce characteristic ergonomic conditions associated with postural loads and movement velocities commonly observed in industrial assembly work. These conditions correspond to a subset of the parameters assessed within the LAL framework, providing quantitative action levels for physical exertion based on postural angles and angular velocities of the neck and upper extremities (Table 2).

The two work tasks were:

1. **Assembly in a non-ideal posture (postural exposure).** This task represented a typical assembly scenario involving sustained postural loading of the upper body. The operator sequentially picked, positioned, tightened, and loosened two triangular metal brackets using a wrench. This procedure involved slow, repetitive forward trunk flexion and prolonged arm elevation while maintaining constrained joint configurations.
2. **Hammering and rapid manipulation (velocity exposure).** This task was designed to evaluate dynamic exposure involving high joint angle velocities and frequent accelerations and decelerations of the upper limbs. The operator executed both hammering and pick-and-place actions, without any predefined order, rhythm or timing. During each hammering instance, strikes were executed with both hands, first with the right and then with the left. This unscripted sequence generated natural variability in direction, amplitude and velocity, enabling the evaluation of each MoCap method's capability to track dynamic and unstructured movements.

### 3.3 Motion-Capture Systems and Data Collection

Two MoCap methods were used to capture joint kinematics during the two work tasks: the IMUs-based technology Xsens [31] and the CV-based method Video Inference for Body Estimation (VIBE) [32]. The two motion-capture systems are shown in Figure 3, illustrating the original video feed, the VIBE 3D reconstructed body mesh, and the Xsens MVN biomechanical model.

The operator wore the Xsens system, which recorded full-body motion through 17 wireless IMUs attached to the main body segments, namely the head, shoulders, sternum, upper and lower arms, hands, pelvis, upper and lower legs, and feet. Tight-fitting clothing was worn over the sensors to secure them and minimise vibrations or displacements while executing the work tasks. It also contributed to a more reliable 3D body mesh reconstruction through VIBE. The sensors measured 3D orientation, acceleration, and angular velocity, enabling real-time tracking of full-body kinematics and postures. Two anthropometric measurements, stature and shoe length, were collected, and the system was calibrated before data collection to align the sensors with body segments. During the calibration procedure, the operator stands for some seconds in N-pose (Figure 3), then starts walking for some seconds and stands again in N-pose.

The VIBE method does not require this calibration procedure. It applies a Deep Learning (DL)-based pose estimation approach to reconstruct 3D posture sequences and joint kinematics from video feed. A single monocular RGB Logitech C922 Pro Stream Webcam recorded the entire task at 60 fps. The camera view captured the entire workstation (Figure 2), providing an unobstructed view of the operator's movements except for occasional self-occlusions inherent to monocular recording. These occlusions pose a significant challenge for accurate joint tracking and illustrate the methodological complexity of single-camera approaches in real industrial conditions.

Each task cycle lasted approximately six minutes and was repeated twice. Recordings from both MoCap systems were conducted simultaneously at 60 Hz to ensure temporal synchronisation and data comparability. All sessions were monitored to verify system operation and sensor placement. Each task was recorded separately to facilitate subsequent processing and analysis.

### 3.4 Data Processing

Both MoCap datasets were processed through parallel workflows (Figure 1) that followed the same overall structure to ensure methodological consistency and enable direct comparison between systems. While Xsens workflow included sensor calibration, the VIBE workflow required

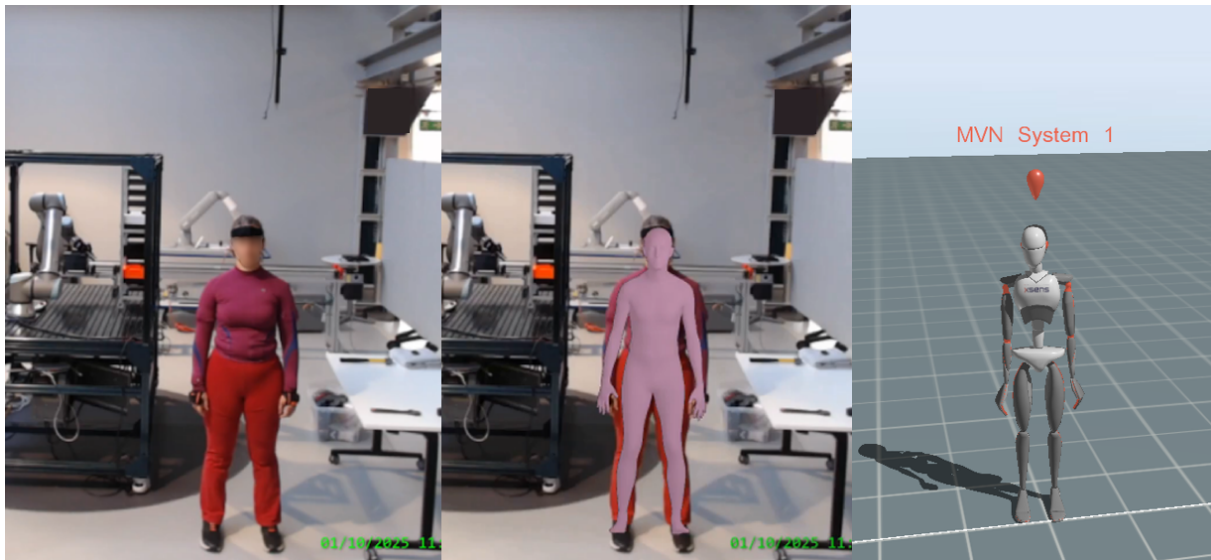


Figure 3: The two motion-capture systems used in the study. Original video feed (left), VIBE 3D reconstructed body mesh (middle), and Xsens MVN Analyze biomechanical model (right).

motion smoothing. All kinematic data were expressed in a common global coordinate system with the vertical axis defined as  $+Z$ , and represented in a pelvis-centred reference frame to standardise spatial alignment across the systems.

Temporal synchronisation was achieved through time-stamped recordings (YYYY-MM-DD HH:MM:SS.sss), ensuring frame-level alignment between Xsens and VIBE. The skeleton models from both systems were hip-locked, i.e. the pelvis was fixed as the reference point to eliminate global translational displacement of the body/skeleton. This established a common spatial origin, allowing segment positions and joint rotations to be compared directly.

Recordings from the Xsens system were processed in MVN Analyze Pro 2024.0 [33] and exported as MVNX files, while the VIBE predictions generated per-frame PKL files that were processed using a Python-based PKL-to-MVNX conversion pipeline developed for this study. This conversion aligned the VIBE output to the Xsens skeleton and segment definitions, reconstructing anatomical segments consistent with the MVN Analyze conventions through the 49-marker used in SPIN evaluation set (joints3d) [34], based on the Human3.6M dataset [35]. As a result, both datasets shared an identical structure, spatial reference frame, and temporal parameters (60 Hz sampling rate).

Two types of data were extracted from each system:

- **Segment position data** to compute upper arm elevation relative to the vertical ( $+Z$ ), defined as the angle between the vector connecting the shoulder and elbow joint centres and the global  $+Z$  axis in accordance with the definition used by Arvidsson et al. [14], where arm elevation was measured with an inclinometer as the angle of the upper arm to the vertical.
- **Joint-angle data** containing local joint rotations to evaluate wrist and neck postures.

To ensure consistent processing, angular velocities were derived from the joint-angle trajectories via finite-difference differentiation. For the Xsens dataset, differentiation was applied

directly to the raw joint angles without any additional filtering, as the Xsens software already incorporates built-in sensor fusion and smoothing algorithms. A linear smoothing technique, the moving average filter [36], was applied to the VIBE dataset for jittering reduction. In this technique, the filtered value is obtained by averaging the signal samples within a defined window that determines the degree of smoothing [36]. A moving-average filter with a 15-frame window (average of 14 preceding frames and the sample, corresponding to 0.25 s at 60 Hz) was applied to the raw joint-angle trajectories prior to differentiation to reduce frame-to-frame jittering inherent to CV-based pose estimation and produced smoother motions [34, 36]. The 15-frame window length was selected based on a 0.25 s time window, commonly used to reduce jittering and to conserve motions in sensor-based human kinematics data filtering [37]. No interpolation or data removal was performed in either dataset to maintain temporal continuity. The resulting datasets were subsequently used to quantify and compare postural and velocity-based exposures.

### 3.5 Ergonomics Evaluation

A musculoskeletal risk assessment method was applied to compare the performance of the CV- and IMUs-based systems. Each system quantified ergonomic risks in regard to posture and motion-related exposures. The assessment method used was the Action Levels for the Prevention of Work-Related Musculoskeletal Disorders in the Neck and Upper Extremities [14]. These action levels are based on research done within occupational and environmental medicine at Lund University, henceforth referred to as LAL. LAL defines action level values (Table 2) indicating increased risk for WMSDs based on joint angular velocities and postural angles, with emphasis on the neck and upper extremities.

Table 2: Proposed action levels for physical workload concerning movement velocities and postures, adapted from [14].

Parameter	10th percentile	50th percentile	90th percentile
<b>Movement velocity</b>			
Upper arm	–	60°/s	–
Wrist (c)	–	20°/s	–
<b>Posture</b>			
Head ext./flex.	–10°	<0° or >25°	50°
Elevated upper arm (d,e)	–	30°	60°

## 4 Data Analysis

Joint kinematics from both motion-capture systems were analysed to compute the parameters defined in the LAL framework. The following quantities were derived for each of the two work tasks:

- Upper-arm elevation, wrist, and neck angles across time
- Angular velocities for the upper arm and wrist obtained from the differentiated joint-angle trajectories.

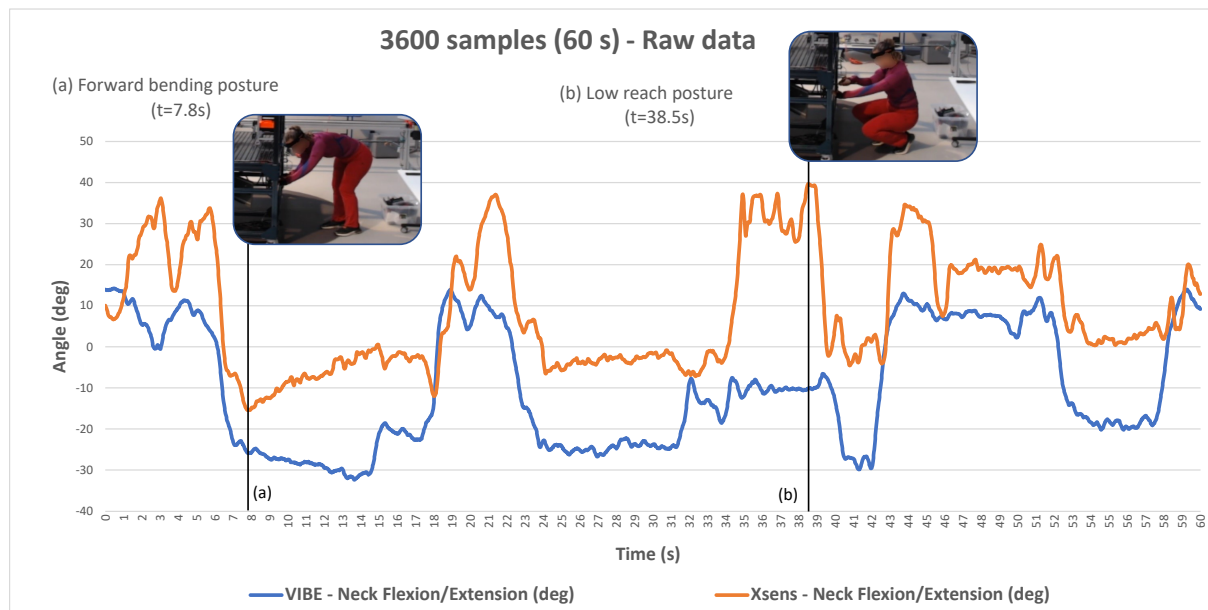


Figure 4: Raw data for neck flexion/extension in Task 1.

#### 4.1 Visual Analysis of Output Data

A visual analysis of the output data obtained from both MoCap systems was performed to examine their temporal correspondence and signal characteristics. This step aimed to verify synchronisation accuracy, detect potential offsets, and identify noise patterns before proceeding with the quantitative analysis based on the metrics defined in the LAL framework.

The data sets for Tasks 1 and 2 were processed as described in Section Data Processing. As stated in the Section Motion-Capture Systems and Data Collection, each recording lasted approximately six minutes. Therefore, a representative 60-second segment (3600 samples at 60Hz) from Task 1 (*Assembly in a Non-ideal Posture*) was selected (Figure 4) to facilitate the visual analysis. This interval captures several forward bending and upward recovery events that represent the characteristics of the postural exposure observed throughout the six-minute task. Figure 4 represents raw neck flexion/extension angles as obtained directly from the two MoCap systems after temporal alignment and coordinate transformation. The horizontal axis represents time in seconds, and the vertical the neck flexion/extension angles in degrees, where the positive values indicate forward flexion and the negative values extension. The orange curve corresponds to Xsens recorded data and the blue curve to VIBE predictions.

Both systems captured the same sequence of forward bending and upright recovery phases, confirming temporal correspondence between datasets. Both datasets are shown unfiltered to reveal their signal characteristics, including smoothness, amplitude, and local noise. Minor jittering is visible in the VIBE signal, typical of frame-to-frame variability in monocular pose estimation [34, 36], while the Xsens trajectory are smoother due to its built-in sensor fusion, which filters noise through combined accelerometer and gyroscope data to stabilise the signal [33]. A consistent negative offset can be observed in the VIBE signal, with angles shifted approximately 15-20° toward flexion compared with Xsens. This difference seems to reflect the absence of a calibration step in the VIBE workflow (Figure 1) that could be the reason for the underestimation of extension and overestimation of flexion relative to Xsens. Despite this offset, the alignment between the signals indicates accurate temporal synchronisation across the two

Table 3: Postural and velocity exposures in Task 1. Cell colours indicate exposure relative to the action levels in Table 2. Green cells represent values within acceptable action levels, and red cells represent values exceeding action levels.

Physical Workload	Xsens			VIBE		
	10%ile	50%ile	90%ile	10%ile	50%ile	90%ile
Head Flexion/Extension (deg)	-4.93	5.95	28.72	-27.92	-16.76	9.53
Upper Arm Angle – Left (deg)	5.46	16.55	37.02	14.21	34.86	49.84
Upper Arm Angle – Right (deg)	4.67	13.65	33.41	13.39	34.93	49.76
Upper Arm Velocity – Left (deg/s)	0.80	5.49	36.47	1.30	7.89	40.75
Upper Arm Velocity – Right (deg/s)	3.46	23.79	73.23	2.11	12.27	40.70
Wrist Velocity – Left (deg/s)	0.55	3.88	39.45	0.60	3.61	12.11
Wrist Velocity – Right (deg/s)	3.08	22.83	85.02	0.51	3.00	10.47

Table 4: Postural and velocity exposures in Task 2. Cell colours indicate exposure relative to the action levels in Table 2. Green cells represent values within acceptable action levels, and red cells represent values exceeding action levels.

Physical Workload	Xsens			VIBE		
	10%ile	50%ile	90%ile	10%ile	50%ile	90%ile
Head Flexion/Extension (deg)	-9.55	11.28	20.76	-3.44	3.91	9.32
Upper Arm Angle – Left (deg)	9.96	24.58	45.28	13.75	32.09	52.70
Upper Arm Angle – Right (deg)	7.27	21.05	44.90	11.32	31.01	63.87
Upper Arm Velocity – Left (deg/s)	5.54	31.27	94.57	4.03	25.67	83.57
Upper Arm Velocity – Right (deg/s)	6.44	40.24	129.00	3.81	25.47	94.81
Wrist Velocity – Left (deg/s)	3.45	23.85	104.91	1.09	7.19	22.51
Wrist Velocity – Right (deg/s)	5.11	35.97	135.57	0.95	5.88	16.68

systems.

These observations establish the baseline correspondence between the two motion-capture systems and highlight their characteristic signal behaviours. The following section presents the quantitative analysis of joint angles and angular velocities, comparing the systems according to the percentile-based metrics defined in the LAL framework.

#### 4.2 Quantitative Analysis

Tables 3 and 4 present the distributions of angular position and velocity recorded by the Xsens system and predicted by the VIBE system for Tasks 1 and 2. These summarise the 10th-, 50th-, and 90th-percentile values for the main joints analysed. Representing typical and peak exposures during each six-minute task cycle. The resulting values were compared against the corresponding action levels proposed by [14] to quantify exposure intensity and identify potential ergonomic risk levels.

The values obtained for Task 1, Table 3, show that both systems captured similar exposure trends, although VIBE predicted systematically higher joint angle values than Xsens, particularly for the upper arms segments. Median upper-arm elevations measured by Xsens were approximately 16–17° on the left and 14° on the right, with 90th-percentile values below 40°, all within the 60° LAL action level indicating low-to-moderate postural load. In contrast, VIBE predicted higher median elevations with angles near 35° and above the 30° action level, and

90th-percentile values around  $50^\circ$  and below the  $60^\circ$  action level for both upper arms. This upward bias suggests an overestimation of elevation likely related to the monocular nature of the vision system, where depth ambiguity, partial self-occlusion, and differences in skeleton definitions (offsets) between the two systems, can distort vertical orientation. At the neck, Xsens recorded neutral postures with a median flexion of approximately  $6^\circ$ , whereas VIBE predicted a median flexion of approximately  $-17^\circ$ . The analysis of the postural exposures seems to indicate a consistent directional offset for posture exposures in LAL. Median upper-arm velocities were modest in both systems  $5\text{--}6^\circ/\text{s}$  for Xsens and  $8\text{--}12^\circ/\text{s}$  for VIBE, remaining well below the  $60^\circ/\text{s}$  action level. Wrist velocities were also low, though the right wrist in Xsens reached a exposure of  $22.8^\circ/\text{s}$ , slightly exceeding the  $20^\circ/\text{s}$  limit. In contrast, VIBE wrist predictions remained within acceptable action levels. Overall, both systems produced consistent exposure patterns, though VIBE yielded higher angular magnitudes and lower velocity amplitudes than Xsens.

In the values obtained for Task 2, Table 4, both systems registered substantially higher velocity exposures and greater variability. Xsens captured median upper-arm elevations of approximately  $24\text{--}25^\circ$  on the left and  $21^\circ$  on the right, with 90th-percentile values below  $46^\circ$ , all within the  $60^\circ$  LAL action level indicating low-to-moderate postural load. In contrast, VIBE predicted higher postural exposures than Xsens, with median elevations with action levels between  $31\text{--}32^\circ$ , above the  $30^\circ$  action level, and 90th-percentile values around  $52\text{--}53^\circ$  for the upper left arm. Notably, the 90th percentile for the right upper arm reached  $63.9^\circ$ , exceeding the  $60^\circ$  action level.

Xsens recorded median upper-arm velocities of  $31\text{--}40^\circ/\text{s}$  remaining below the  $60^\circ/\text{s}$  action level, and 90th-percentile peaks of  $95\text{--}129^\circ/\text{s}$ . Wrist velocities were also higher  $23\text{--}36^\circ/\text{s}$ , exceeding the  $20^\circ/\text{s}$  LAL action level and indicating high-velocity exposure. VIBE followed the same exposure trend but reported lower amplitudes for the median upper-arm velocities of around  $25\text{--}26^\circ/\text{s}$ , and the 90th-percentile peaks ( $83\text{--}95^\circ/\text{s}$ ). Wrist median velocities of  $5\text{--}8^\circ/\text{s}$  and 90th-percentile peaks of  $16\text{--}23^\circ/\text{s}$ , highlighted system differences. These discrepancies likely reflect the smoothing effects of the 15-frame moving-average filter and a minor underestimation during rapid transitions.

Overall, Xsens and VIBE exhibited consistent exposure trends, although systematic amplitude and velocity differences were observed between the two systems. Despite these systematic offsets, both MoCap systems produced the same ergonomic classifications under the LAL framework: Task 1 as low-to-moderate risk and Task 2 as high-velocity, high-exposure.

## 5 Discussion

The comparative analysis between the IMUs-based Xsens and the CV-based VIBE systems revealed systematic but comparable differences in both postural and velocity metrics that are consistent with the visual analysis in Figure 4. VIBE produced slightly higher postural angles and lower velocity amplitudes than Xsens, suggesting a bias linked to the absence of a calibration step and to the temporal damping introduced by the 15-frame moving-average filter. VIBE showed a mean offset across both work tasks, approximately  $15\text{--}20^\circ$  in neck flexion and  $10\text{--}20^\circ$  in upper-arm elevations, while maintaining high temporal correspondence. These findings indicate that despite absolute magnitude deviations, VIBE predicted exposure patterns and task-specific temporal dynamics accurately.

This behaviour aligns with previous comparative studies of optical and inertial systems. Meletani et al. [21] similarly reported strong agreement between Move4D stereophotogrammetry and Xsens for gait parameters, but observed moderate discrepancies in absolute joint angles. Likewise, Scataglini et al. [23] found that camera-based systems present good concurrent validity in sagittal-plane kinematics but reduced accuracy in transverse and frontal planes.

The consistent overestimation of flexion in VIBE mirrors the depth-ambiguity and occlusion sensitivities reported in these optical studies.

The underestimation of velocity amplitudes by VIBE is also coherent with evidence from Das et al. [28], who observed systematic amplitude damping in MCBS due to temporal smoothing and frame-rate constraints. The applied moving-average filter in this work successfully reduced jittering but attenuated fast transitions, explaining the lower 90th-percentile velocity values obtained compared to Xsens. Despite these differences, both systems classified Task 1 as low-to-moderate risk and Task 2 as high-velocity, high-exposure under the LAL framework, confirming that relative exposure ranking remains robust even when amplitude scaling differs.

From an ergonomics-application standpoint, these findings extend prior observations [26, 25, 29] that showed that CV-based methods can replicate posture trends and enable continuous workplace monitoring but remain limited by calibration, lighting, and environmental variability. The consistent exposure trends obtained across both tasks support those conclusions and demonstrate that monocular CV-based MoCap can capture representative ergonomic indicators in realistic industrial settings. However, the observed negative bias in neck extension and overestimation in arm elevation reinforce that monocular configurations are best suited for comparative or relative analyses rather than for absolute kinematic quantification. At the same time, the results corroborate earlier findings [19, 27], highlighting the strengths of IMUs-based systems in stability, temporal fidelity, and independence from external lighting. Xsens maintained higher sensitivity to rapid wrist motions and asymmetrical loads, confirming its suitability for high-precision or certification-grade assessments. Nevertheless, its intrusive nature and calibration requirements indicate the value of CV-based systems as complementary, low-cost tools for large-scale monitoring, consistent with the hybrid or multi-sensor strategies advocated in the literature [26, 19].

The quantitative comparison presented here bridges the gap between laboratory-based gait validations and real industrial ergonomics applications, providing empirical evidence that single-camera approaches—though not yet a replacement for IMU systems—are mature enough to support continuous, non-intrusive monitoring within Industry 5.0's human-centred paradigm.

Overall, this comparative analysis demonstrates that DL-based CV methods, such as VIBE, can provide representative and interpretable motion data for direct-measurement ergonomics evaluations when appropriately filtered and benchmarked. The quantitative comparison presented here bridges the gap between laboratory-based gait validations and real industrial ergonomics applications, providing empirical evidence that even though single-camera approaches are not yet a replacement for IMU systems, they are mature enough to support continuous, non-intrusive monitoring within the Industry 5.0's human-centred paradigm.

## 6 Conclusion and Future Work

This study compared a CV-based (VIBE) and an IMUs-based (Xsens) MoCap systems using the LAL musculoskeletal risk assessment method to quantify work postures and movement velocities exposures in two representative industrial assembly tasks. Both systems produced consistent ergonomic classifications, Task 1 (Assembly in Non-Ideal Posture) as low-to-moderate risk and Task 2 (Hammering and Rapid Manipulation) as high-velocity, high-exposure. These findings demonstrate that CV-based approaches can generate representative and interpretable kinematic data for direct-measurement ergonomics methods when appropriate filtering and benchmarking are applied. The findings validate that CV-based systems can provide representative motion data for direct measurement ergonomics evaluations when appropriate filtering and alignment are applied. The study extends prior work by demonstrating the feasibility of monocular RGB-based pose estimation in a real industrial setting, bridging the performance gap between research-grade

IMUs systems and accessible, non-intrusive CV solutions.

Future work will extend this comparative framework in several directions. From a methodological standpoint, a calibration procedure for camera-based systems will be developed to minimise baseline offsets and improve cross-system comparability. The next research phase will also introduce force-based ergonomics evaluation methods, such as the Arm Force Field (AFF) [38], to complement posture- and velocity-based indicators and enable more comprehensive assessments of physical workload. Technically, integrating MoCap outputs with digital human modelling (DHM) tools will be pursued to allow combined analyses of human motion, forces, and environmental interaction using detailed workstation models (e.g. CAD layouts). Finally, testing the proposed approach in a real industrial case study will be essential to validate its robustness and scalability under authentic production conditions.

### Acknowledgements

The authors acknowledge the financial support received from KK-Stiftelsen (The Knowledge Foundation, Stockholm, Sweden) through the Synergy programme for the research project *LITMUS: Leveraging Industry 4.0 Technologies for Human-Centric Sustainable Production* (grant #2024-0013), and from the research profile *VF-KDO: Virtual factories with knowledge-driven optimization* at the University of Skövde (grant #2018-0011), and from Vinnova, Sweden's Innovation Agency, through the Advanced Digitalization programme, for the project *AI Support and Digital Human Models for Time Data Management: TIMEBLY 2* (grant #2025-01012).

### References

- [1] Lasi H, Fettke P, Kemper HG, Feld T, Hoffmann M. Industry 4.0. *Business & Information Systems Engineering*. 2014 Aug;6(4):239-42.
- [2] Liao Y, Deschamps F, Loures EdFR, Ramos LFP. Past, present and future of Industry 4.0 - a systematic literature review and research agenda proposal. *International Journal of Production Research*. 2017 Jun;55(12):3609-29. Publisher: Taylor & Francis.
- [3] Xu X, Lu Y, Vogel-Heuser B, Wang L. Industry 4.0 and Industry 5.0—Inception, conception and perception. *Journal of Manufacturing Systems*. 2021 Oct;61:530-5.
- [4] Sun S, Zheng X, Villalba-Díez J, Ordieres-Meré J. Data Handling in Industry 4.0: Interoperability Based on Distributed Ledger Technology. *Sensors*. 2020 Jan;20(11):3046. Publisher: Multidisciplinary Digital Publishing Institute.
- [5] Dall'Ora N, Alamin K, Fraccaroli E, Poncino M, Quaglia D, Vinco S. Digital Transformation of a Production Line: Network Design, Online Data Collection and Energy Monitoring. *IEEE Transactions on Emerging Topics in Computing*. 2022 Jan;10(1):46-59.
- [6] Kusiak A. Open manufacturing: a design-for-resilience approach. *International Journal of Production Research*. 2020 Aug;58(15):4647-58. Publisher: Taylor & Francis.
- [7] Grosse EH, Sgarbossa F, Berlin C, Neumann WP. Human-centric production and logistics system design and management: transitioning from Industry 4.0 to Industry 5.0. *International Journal of Production Research*. 2023 Nov;61(22):7749-59. Publisher: Taylor & Francis.
- [8] Industry 5.0: towards a sustainable, human centric and resilient European industry. Publications Office of the European Union; 2021.
- [9] Mourtzis D, Angelopoulos J, Panopoulos N. Operator 5.0: A Survey on Enabling Technologies and a Framework for Digital Manufacturing Based on Extended Reality. *Journal of Machine Engineering*. 2022 Mar;22(1):43-69. Publisher: Editorial Institution of Wrocław Board of Scientific Technical Societies Federation NOT.

- [10] Hignett S, McAtamney L. Rapid Entire Body Assessment (REBA). *Applied Ergonomics*. 2000 Apr;31(2):201-5.
- [11] McAtamney L, Nigel Corlett E. RULA: a survey method for the investigation of work-related upper limb disorders. *Applied Ergonomics*. 1993 Apr;24(2):91-9.
- [12] Rhén IM, Forsman M. Inter- and intra-rater reliability of the OCRA checklist method in video-recorded manual work tasks. *Applied Ergonomics*. 2020 Apr;84:103025.
- [13] Takala EP, Pehkonen I, Forsman M, Hansson G Mathiassen SE, Neumann WP, et al. Systematic evaluation of observational methods assessing biomechanical exposures at work. *Scandinavian Journal of Work, Environment and Health*. 2010;36(1):3-24. Publisher: Finnish Institute of Occupational Health.
- [14] Arvidsson I, Dahlqvist C, Enquist H, Nordander C. Action Levels for the Prevention of Work-Related Musculoskeletal Disorders in the Neck and Upper Extremities: A Proposal. *Annals of Work Exposures and Health*. 2021 Aug;65(7):741-7. Publisher: Oxford University Press (OUP).
- [15] Hansson G Balogh I, Ohlsson K, Granqvist L, Nordander C, Arvidsson I, et al. Physical workload in various types of work: Part I. Wrist and forearm. *International Journal of Industrial Ergonomics*. 2009 Jan;39(1):221-33.
- [16] Bhatia V, Vaishya RO, Jain A, Grover V, Arora S, Das G, et al. Static and dynamic validation of kinect for ergonomic postural analysis using electro-goniometers as a gold standard:A preliminary study. *Technology and Health Care*. 2023 Nov;31(6):2107-23. Publisher: SAGE Publications.
- [17] Iriondo Pascual A, Holm M, Ng A, Larsson F, Olsson J. Integrating Motion Capture and Digital Human Modelling Tools for Evaluating Worker Ergonomics - A Case Study in a Medium Size Enterprise Assembly Station. In: Kurosu M, Hashizume A, editors. *Human-Computer Interaction*. Cham: Springer Nature Switzerland; 2025. p. 362-73.
- [18] Fontinovo E, Luque EP, Papetti A, Högberg D, Hanson L, Truijen S, et al. Comparison Between Observational Method, Wearable Inertial Measurement System and 4D Stereophotogrammetry for Ergonomics Risk Assessment: A Case Study. In: Marshall R, Summerskill S, Harih G, Scataglini S, editors. *Advances in Digital Human Modeling II*. Cham: Springer Nature Switzerland; 2025. p. 193-206.
- [19] Ciccarelli M, Papetti A, Germani M. Empowering industry 5.0: automated sensor-based ergonomic risk assessment. *International Journal on Interactive Design and Manufacturing (IJIDeM)*. 2025 Nov;19(11):7731-53.
- [20] Plantard P, Shum HPH, Le Pierres AS, Multon F. Validation of an ergonomic assessment method using Kinect data in real workplace conditions. *Applied Ergonomics*. 2017 Nov;65:562-9.
- [21] Meletani S, Scataglini S, Mandolini M, Scalise L, Truijen S. Experimental Comparison between 4D Stereophotogrammetry and Inertial Measurement Unit Systems for Gait Spatiotemporal Parameters and Joint Kinematics. *Sensors*. 2024 Jan;24(14):4669. Number: 14 Publisher: Multidisciplinary Digital Publishing Institute.
- [22] Cao Z, Hidalgo G, Simon T, Wei SE, Sheikh Y. OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. *IEEE transactions on pattern analysis and machine intelligence*. 2021 Jan;43(1):172-86.
- [23] Scataglini S, Abts E, Van Bocxlaer C, Van den Bussche M, Meletani S, Truijen S. Accuracy, Validity, and Reliability of Markerless Camera-Based 3D Motion Capture Systems versus Marker-Based 3D Motion Capture Systems in Gait Analysis: A Systematic Review and Meta-Analysis. *Sensors*. 2024 Jan;24(11):3686. Number: 11 Publisher: Multidisciplinary Digital Publishing Institute.
- [24] Ceseracciu E, Sawacha Z, Cobelli C. Comparison of Markerless and Marker-Based Motion Capture Technologies through Simultaneous Data Collection during Gait: Proof of Concept. *PLOS ONE*. 2014 Mar;9(3):e87640. Publisher: Public Library of Science.

- [25] Elango V, Hedelin S, Hanson L, Sandblad J, Syberfeldt A, Forsman M. Evaluating ERAIVA - a software for video-based awkward posture identification. *International Journal of Human Factors and Ergonomics*. 2024 Jan;11(6):1-16. Publisher: Inderscience Publishers.
- [26] Elango V, Petravic S, Hanson L. Evaluation of upper body postural assessment of forklift driving using a single depth camera. *Proceedings of the 7th International Digital Human Modeling Symposium*. 2022 Aug;7(1). Publisher: University of Iowa.
- [27] Lind CM, Sandsjö L, Mahdavian N, Högberg D, Hanson L, Olivares JAD, et al. Prevention of Work-Related Musculoskeletal Disorders Using Smart Workwear – The Smart Workwear Consortium. In: Ahram T, Karwowski W, Taiar R, editors. *Human Systems Engineering and Design*. Cham: Springer International Publishing; 2019. p. 477-83.
- [28] Das K, de Paula Oliveira T, Newell J. Comparison of markerless and marker-based motion capture systems using 95% functional limits of agreement in a linear mixed-effects modelling framework. *Scientific Reports*. 2023 Dec;13(1):22880. Publisher: Nature Publishing Group.
- [29] Agostinelli T, Generosi A, Ceccacci S, Mengoni M. Validation of computer vision-based ergonomic risk assessment tools for real manufacturing environments. *Scientific Reports*. 2024 Nov;14(1):27785. Publisher: Nature Publishing Group.
- [30] ASSAR Labs;. Available from: <https://assarinnovation.se/assar-labs/>.
- [31] Xsens Awinda | Movella.com;. [cited 2025 Oct 18. Available from: <https://www.movella.com/motion-capture/xsens-mvn-awinda>.
- [32] Kocabas M, Athanasiou N, Black MJ. VIBE: Video Inference for Human Body Pose and Shape Estimation. In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; 2020. p. 5252-62. ISSN: 2575-7075.
- [33] Xsens Analyze | Movella.com;. [cited 2025 Oct 18]. Available from: <https://www.movella.com/motion-capture/mvn-analyze>.
- [34] Kolotouros N, Pavlakos G, Black MJ, Daniilidis K. Learning to Reconstruct 3D Human Pose and Shape via Model-fitting in the Loop. *arXiv*; 2019. ArXiv:1909.12828 [cs].
- [35] Ionescu C, Papava D, Olaru V, Sminchisescu C. Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2014 Jul;36(7):1325-39.
- [36] Martini E, Calanca A, Bombieri N. Denoising and completion filters for human motion software: A survey with code. *Computer Science Review*. 2025 Nov;58:100780.
- [37] Rhen IM, Gyllensvärd D, Hanson L, Högberg D. Time dependent exposure analysis and risk assessment of a manikin's wrist movements. *Université Claude Bernard Lyon*; 2011. .
- [38] La Delfa NJ, Potvin JR. The 'Arm Force Field' method to predict manual arm strength based on only hand location and force direction. *Applied Ergonomics*. 2017 Mar;59:410-21.