



**HAL**  
open science

## **Human Movement and Ergonomics: an Industry-Oriented Dataset for Collaborative Robotics**

Pauline Maurice, Adrien Malaisé, Clélie Amiot, Nicolas Paris, Guy-Junior Richard, Olivier Rochel, Serena Ivaldi

► **To cite this version:**

Pauline Maurice, Adrien Malaisé, Clélie Amiot, Nicolas Paris, Guy-Junior Richard, et al.. Human Movement and Ergonomics: an Industry-Oriented Dataset for Collaborative Robotics. The International Journal of Robotics Research, 2019, 38 (14), pp.1529-1537. 10.1177/0278364919882089 . hal-02289107

**HAL Id: hal-02289107**

**<https://hal.science/hal-02289107v1>**

Submitted on 16 Sep 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Human Movement and Ergonomics: an Industry-Oriented Dataset for Collaborative Robotics

Journal Title  
XX(X):1-8  
©The Author(s) 2019  
Reprints and permission:  
sagepub.co.uk/journalsPermissions.nav  
DOI: 10.1177/ToBeAssigned  
www.sagepub.com/

SAGE

Pauline Maurice<sup>1</sup>, Adrien Malaisé<sup>1</sup>, Clélie Amiot<sup>1</sup>, Nicolas Paris<sup>1</sup>, Guy-Junior Richard<sup>1</sup>, Olivier Rochel<sup>1</sup> and Serena Ivaldi<sup>1</sup>

## Abstract

Improving work conditions in industry is a major challenge that can be addressed with new emerging technologies such as collaborative robots. Machine learning techniques can improve the performance of those robots, by endowing them with a degree of awareness of the human state and ergonomics condition. The availability of appropriate datasets to learn models and test prediction and control algorithms however remains an issue. This paper presents a dataset of human motions in industry-like activities, fully labeled according to the ergonomics assessment worksheet EAWS, widely used in industries such as car manufacturing. Thirteen participants performed several series of activities, such as screwing and manipulating loads in different conditions, resulting in more than 5 hours of data. The dataset contains the participants' whole-body kinematics recorded both with wearable inertial sensors and marker-based optical motion capture, finger pressure force, video recordings, and annotations by 3 independent annotators of the performed action and the adopted posture following the EAWS postural grid. Sensor data are available in different formats to facilitate their reuse. The dataset is intended for use by researchers developing algorithms for classifying, predicting or evaluating human motion in industrial settings, as well as researchers developing collaborative robotics solutions that aim at improving the workers' ergonomics. The annotation of the whole dataset following an ergonomics standard makes it valuable for ergonomics-related applications, but we expect its use to be broader in the robotics, machine learning and human movement communities.

## Keywords

Human motion dataset, Industrial activities, Optical motion capture, Inertial motion capture, Ergonomics, Posture and action annotation.

## 1 Introduction

Work-related musculoskeletal disorders (WMSDs) are among the first causes of occupational diseases in many countries worldwide. They represent a major health issue and an important cost for companies and society (Schneider et al. 2010; US Bureau of Labor Statistics 2016). WMSDs develop when biomechanical demands at work repeatedly exceed the worker's physical capacity, which happens due to force exertion but also extreme postures (Punnett and Wegman 2004). Improving work conditions in industry is therefore a major challenge that can be addressed with new emerging technologies such as collaborative robots to assist workers (Marin et al. 2018; Kim et al. 2018a; Gopinathan et al. 2018; Busch et al. 2018; Pearce et al. 2018; El Makrini et al. 2019), or wearable sensors and devices to monitor and warn of health-threatening situations (Vignais et al. 2013; Yan et al. 2017; Kim et al. 2018b).

Benefiting from machine learning techniques, the efficacy of such assistive and warning devices can be further improved by endowing them with a degree of awareness of the human state and ergonomics condition. For instance, Coupeté et al. (2018) and Unhelkar et al. (2018) perform human motion recognition and prediction respectively to plan the behavior of a collaborative robot that enables efficient and safe interaction in industrial applications. Malaisé et al. (2019) focus on ergonomics and propose

an activity recognition module that automatically identifies non-ergonomic gestures during industry-oriented activities. Marin et al. (2018) learn surrogate models to compute in real-time the most ergonomic posture for a drilling task and adapt the motion of a collaborative robot accordingly. Pernel et al. (2018) also use surrogate models to estimate muscle fatigue in a collaborative human-robot polishing task.

Machine learning algorithms, however, often require training data whose acquisition is time-consuming and necessitates specific equipment. Alternatively, algorithms can be trained and tested using publicly available human motion datasets. But while numerous motion datasets are available, they largely focus on daily life activities or sport movements (Mandery et al. 2015; CMU 2019; Kuehne et al. 2011; Xia et al. 2012; Guerra-Filho and Biswas 2012; Ruffieux et al. 2015). Those datasets are therefore not adequate for learning models of industrial tasks. Mandery et al. (2015) and CMU (2019) created comprehensive motion capture datasets which include trials of industry-oriented activities (e.g., manipulating a screw

<sup>1</sup>Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France

## Corresponding author:

Pauline Maurice, INRIA Nancy Grand-Est, 615 rue du jardin botanique, 54600 Villers-lès-Nancy, France.

Email: pauline.maurice@polytechnique.org

driver or hammer, lifting loads). Those industry-oriented motions are however very few in both datasets, mostly performed by a single subject and not repeated. Hence movement and posture variability cannot be accounted for, which limits the generalization ability of models of motion (Perez and Nussbaum 2006; Gaudes et al. 2016). Recently, Shah and colleagues created a human motion dataset of industrial activities to test their algorithms for human-robot collaboration (Hayes and Shah 2017; Kubota et al. 2019). However, only arm motions were recorded. Demonstrations of whole-body tasks and postures typically observed on assembly lines and evaluated in standard ergonomics assessment tools are therefore not available (David 2005). Furthermore, supervised learning –which represents a large part of machine learning techniques– requires labeled training data. Most existing human motion datasets include annotations of the activity being performed, but to the best of our knowledge, none of them include ergonomics-oriented labels.

This paper describes the creation of a dataset that contains industry-oriented motions with labels relevant for ergonomics studies, human movement studies, and model learning for collaborative robotics. Aside from its unique activities and labels, our dataset contains kinematic data from both wearable and external motion capture sensors, as well as videos, in order to facilitate the reuse of the data by various communities and for various applications. We created this dataset to advance our research in ergonomics-aware control of collaborative robots and prediction of whole-body movements, in the context of the European Project An.Dy (Ivaldi et al. 2017), where we envision collaborative robots and wearable sensors to improve the ergonomic conditions of workers in factories. But due to the growing interest in the latter topic on one hand, and the time and effort needed to create and annotate a dataset on the other hand, we envision that our human motion dataset will be of great interest for the broader research community.

The paper is organized as follows. Section 2 describes the data collection campaign, *i.e.*, the activities recorded, the sensors used and the data collected, and the annotation process. Section 3 presents the organization of the dataset and evaluates the quality of the data. Finally, Section 4 concludes with closing remarks and future directions.

## 2 Methods

Awkward postures represent one of the major WMSDs risk factor (Punnett and Wegman 2004). Consequently, most standard ergonomics assessment methods used in industry focus on postural evaluation (David 2005) (*e.g.*, the Ergonomic Assessment Worksheet (EAWS) (Schaub et al. 2013), the Rapid Upper Limb Assessment (RULA) (McAtamney and Corlett 1993), the Rapid Entire Body Assessment (REBA) (Hignett and McAtamney 2000), the Owako Working Posture Analysis System (OWAS) (Karhu et al. 1977)). Our dataset focuses on industry-oriented activities in which participants adopt a variety of postures typically evaluated in the aforementioned ergonomics worksheets. Note that the data were collected in a lab environment, and not in a real factory. Lab recording allows a ground truth measurement of the human kinematics with

a marker-based optical motion capture system, which can hardly be used in a cluttered industrial setting. This section describes the process for data collection and annotation.

### 2.1 Participants

13 healthy adults participated in the data collection (9 males, 4 females). Their average age was 25.7 yrs (SD = 5.0 yrs), their average stature was 175.4 cm (SD = 7.9 cm), and their average body mass was 72.3 kg (SD = 14.4 kg). Participants were students and researchers with no or limited experience of industrial work. The data collection was approved by INRIA’s ethical committee (COERLE). All participants gave written informed consent before starting the data collection. Participants were assigned a numerical ID (random number between 0 and 10000) which served to anonymize the data.

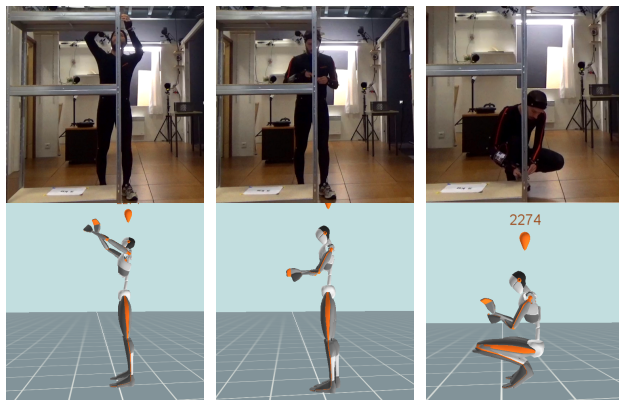
### 2.2 Experimental Set-up

The dataset targets postures and actions that are commonly observed in industrial settings such as assembly lines, and that are evaluated in standard ergonomics assessments (*e.g.*, load manipulation, overhead work). Participants performed a series of 6 industry-oriented activities (Fig. 1), designed with the help of an industrial ergonomist and inspired by a car manufacturing use case:

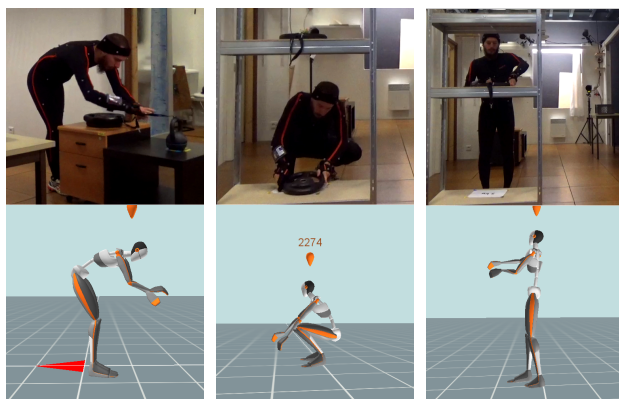
- *Screw high* (SH): Take a screw and a bolt on a 75 cm-high table, walk to the shelf, screw at a height of 175 cm (performed with bare hands, *i.e.*, no screwdriver).
- *Screw middle* (SM): Take a screw and a bolt on a 75 cm-high table, walk to the shelf, screw at a height of 115 cm.
- *Screw low* (SL): Take a screw and a bolt on a 75 cm-high table, walk to the shelf, screw at a height of 25 cm (6 participants) or 60 cm (7 participants).
- *Untie knot* (UK): Untie a knot placed on a 45 cm-high table.
- *Carry 5 kg* (C5): Take a 5 kg load on a 55 cm-high table, walk to the shelf, put the load on a 20 cm-high shelf.
- *Carry 10 kg* (C10): Take a 10 kg load on a 55 cm-high table, walk to the shelf, put the load on a 110 cm-high shelf.

Items used in the activities were positioned to encourage the adoption of specific postures (*e.g.*, raised arms, bent torso, crouching). Participants were however not given any instructions regarding the way to perform each activity, *i.e.*, they were not verbally constrained to adopt a specific posture.

One trial consisted in performing all 6 activities successively in a given order, selected among 6 predefined sequences (Seq. 1: SL-SM-SH-UK-C10-C5; Seq. 2: SH-SM-SL-UK-C10-C5; Seq. 3: SL-SM-UK-C10-SH-C5; Seq. 4: SL-UK-C10-SM-C5-SH; Seq. 5: UK-C10-SH-SM-C5-SL; Seq. 6: UK-C10-C5-SH-SM-SL). Each participant performed 3 different sequences randomly chosen among the 6 possible ones, with 5 consecutive trials for each sequence, resulting in a total of 15 trials per participant. Each trial lasted approximately 90 s, and a break of about 2 min was allowed between each trial to limit fatigue. Before starting



(a) Screw high (SH) (b) Screw middle (SM) (c) Screw low (SL)



(d) Untie knot (UK) (e) Carry 5kg (C5) (f) Carry 10kg (C10)

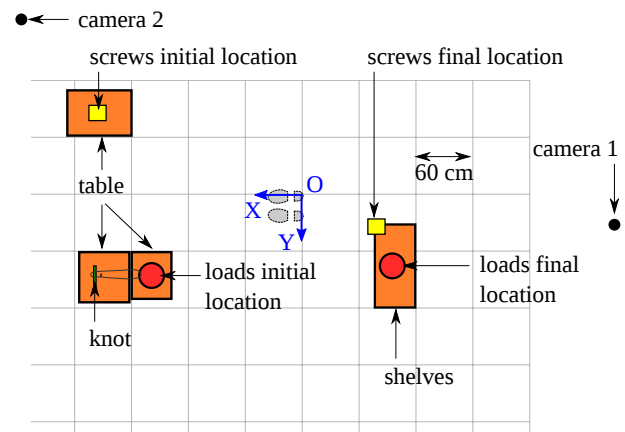
**Figure 1.** Picture and Xsens avatar view of a participant performing the 6 activities included in the data collection.

the data collection, participants practiced each activity a couple of times. Participants performed their trials using one of the 2 spatial organizations depicted in Fig. 2 (6 participants for organization A and 7 for organization B). Participants started and ended each trial in a fixed position: standing arms along the body, the rear of the right foot on the origin of the motion capture world frame, with X axis facing forward (see Fig. 2).

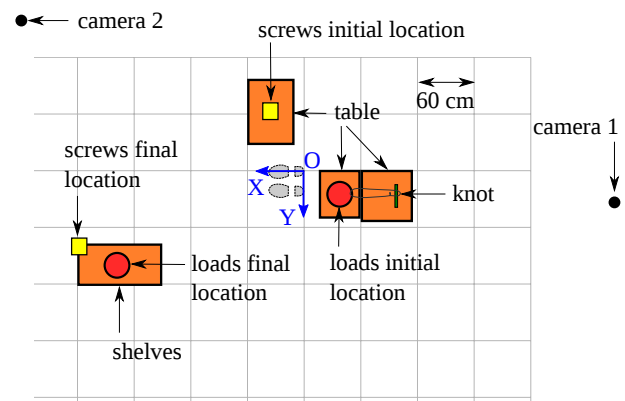
The dataset therefore includes a diversity of participants, as well as several series of activities performed in different orders and with different set-ups, plus several repetitions of a same series for each participant, in order to account for inter and intra-individual movement variability.

### 2.3 Instrumentation

In human movement science, movement kinematics is traditionally recorded with specific motion capture systems. Among all existing modalities, marker-based optical motion capture remains the gold standard method to record human kinematics, mostly because of its accuracy. But recently, inertial motion capture using wearable sensors (IMU-based) has received a lot of attention. Wearable sensors are specifically interesting in cluttered environments –such as factories– where optical motion capture markers are easily occluded and reflective surfaces may perturb the measures. Image-based motion analysis is also undergoing a rapid development, and libraries such as OpenPose now enable



(a) Set-up A



(b) Set-up B

**Figure 2.** Spatial organization of the items used during the data collection. 6 participants used set-up A and 7 used set-up B. The blue frame corresponds to the motion capture world frame (Qualisys and Xsens). The grey feet silhouettes represent the participant's initial and final position in each trial. The positions of the tables and shelves are exact, whereas the positions of the 2 cameras are approximative.

reliable human pose detection from regular videos (Cao et al. 2018). Though less accurate than marker-based motion capture and equally affected by occlusions, video recording has the advantage of being cheaper and less invasive than systems requiring wearable sensors or markers\*. Since those 3 modalities – optical motion capture, inertial motion capture, and video– are in use in the human-robot interaction community, they were all included in the dataset. In addition to whole-body kinematic data, we captured finger forces and motions using a sensorized glove.

**2.3.1 Inertial motion capture:** Participants were equipped with the Xsens MVN Link system<sup>†</sup> (Xsens, Enschede, The Netherlands) to record whole-body kinematics. The MVN Link system consists of 17 IMUs placed all over the body

\*It is, however, often difficult to track the entire human body in factories using only video cameras, because of the high number of required cameras and the occlusions due to the working environment. Cameras may therefore be effective for tracking human movements across the workspace, but not for a precise estimation of the body posture –which is necessary for an accurate ergonomics evaluation.

<sup>†</sup><https://www.xsens.com>



to measure the orientation of the body segments: 1 sensor on the head, 1 on the sternum, 1 on the pelvis, 1 on each scapula, 1 on each upper arm, 1 on each forearm, 1 on each hand, 1 on each thigh, 1 on each shank and 1 on each foot. Xsens MVN whole-body lycra suits of different sizes (M to XXL) were used to ensure correct positioning of all sensors. The Xsens system was calibrated with the “N-pose and walk” calibration procedure every 5 trials, for each participant. The Xsens world frame was reinitialized at the beginning of each trial to match the Qualisys world frame. Xsens data were recorded at 240 Hz (recommended MVN Link capture rate) using the Xsens MVN Analyze software (version 2018.0.0).

**2.3.2 Optical motion capture:** A Qualisys motion capture system<sup>‡</sup> (Qualisys, Göteborg, Sweden) was used to record whole-body kinematics. 12 Oqus cameras were positioned all over the workspace. Participants were equipped with 43 spherical reflective markers of diameter 12.5 mm. 39 markers were positioned on the body according to the Vicon Plug-In Gait Marker Placement recommendations (the sacral marker was not used) (Vicon 2017). Two additional markers were placed on each foot: one on the 1<sup>st</sup> metatarsal head, on the midfoot side of the equinus break between forefoot and midfoot (labeled LTOA and RTOA for the left and right foot respectively), and one on the 5<sup>th</sup> metatarsal head, on the midfoot side of the equinus break between forefoot and midfoot (labeled LTOB and RTOB). A calibration of the Qualisys system was performed at the beginning of the recording session for each participant. Qualisys data were recorded at 120 Hz (sub-multiple of Xsens capture rate) with the Qualisys Track Manager software (version 2.16). After the recording, markers were manually labeled by one of the experimenter in the QTM software, and the labeling was double-checked by a second experimenter.

**2.3.3 Hand contact and finger flexion:** Participants were equipped with a prototype e-glove from Emphasis Telematics<sup>§</sup> (Emphasis Telematics, Athens, Greece) on their right hand (regardless of their handedness). The glove embeds 3 flexion sensors –on the thumb, index, and middle finger– and 4 pressure sensors –on the palm and on the fingertip of the thumb, index and middle finger. The flexion sensors measure the angle between the orientation of the palm and of the last phalange of the finger. The e-glove communicates with its server running on a computer via Wi-Fi. The e-glove internal time was synchronized with the computer Unix time before each recording session, using a synchronization function of the e-glove software. Data of the e-glove were recorded at 50 Hz with a proprietary software.

**2.3.4 Video recording:** The data collection was recorded with 2 video cameras. Approximate positions of the cameras are displayed in Fig. 2. The positions of the cameras were not calibrated with the Qualisys motion capture system, however the grid formed by the floor tiles could be used to calibrate the cameras post-hoc (tiles dimension: 60 cm). Videos were recorded at 25 fps. A custom-written script based on the OpenPose library (Cao et al. 2018) was used in a post-processing phase to blur participants’ faces and anonymize the videos.

**2.3.5 Synchronization:** A custom-written script was used to launch and stop the recording of all sensors (all acquisition

software were running on the same machine). Proper synchronization was achieved by time-stamping the data of all sensors, except the video cameras. Xsens and e-glove software use Unix timestamps, while Qualisys uses machine-specific timestamps. The correspondence between the Unix time and the machine time was therefore logged at the beginning of each trial, and Qualisys timestamps were converted into Unix timestamps in a post-processing phase. Conversely, video cameras were used separately and could not be automatically synchronized with the other sensors.

## 2.4 Annotations

After the data collection, 3 independent human annotators manually labeled the motions in all trials, using a taxonomy of actions and postures that we defined beforehand (Table 1). For each trial, the 3 annotators were randomly selected among a pool of 4 annotators. We based the labels of our taxonomy on the EAWS worksheet, that we selected among all ergonomics assessment worksheets, because it is widely used in industry (Schaub et al. 2013). The taxonomy contains 3 levels of annotations: *general posture*, *detailed posture*, and *current action*, detailed in Table 1. The *general posture* level describes the whole-body posture, mainly following the categories of postures listed in the EAWS worksheet (e.g., standing, kneeling, lying). Note that *lying* and *sitting* are not present in the dataset, but are included in the taxonomy for comprehensiveness with respect to the EAWS worksheet. The *detailed posture* level combines the general posture with the position of the trunk and arms, following EAWS sub-categories (e.g., bent forward, hand above head level). The *current action* level corresponds to goal-oriented actions, e.g., carrying a load, screwing, reaching. While the *general posture* and *detailed posture* labels can cover any motion, the list of *current action* labels contains only actions that are part of the activities performed in the dataset.

Annotations were performed with the Anvil software<sup>¶</sup>, on the videos of the Xsens avatar replaying the participants’ motions, exported from the MVN Analyze software. Each frame of the video was given 3 different labels, one for each level of the taxonomy (video framerate: 24 fps). Annotators could also consult the videos from the 2 cameras to help resolve potential labeling ambiguity in the Xsens generated videos (videos from the cameras include helpful contextual information, whereas Xsens videos only display the avatar’s motion).

## 3 Dataset

The dataset containing the data from all the sensors and in several formats is available on Zenodo: <https://zenodo.org/record/3254403>. The amount of data corresponds to approximately 5 hours of recording (13 participants, 15 trials per participant, about 90 s per trial). This section describes the organization of the dataset, and evaluates the quality and reliability of the data.

<sup>‡</sup><https://www.qualisys.com>

<sup>§</sup><http://www.emphasisnet.gr/e-glove/#1476464465174-6fa88672-0410>

<sup>¶</sup><http://www.anvil-software.org/>

Label	State	Description
<i>general posture</i>		Main posture
St	Standing	Ends when feet start moving.
Wa	Walking	Starts when one foot start moving, ends when both feet are still.
Kn	Kneeling	At least one knee on the floor.
Cr	Crouching	No knee on the floor.
Si	Sitting	Buttock on a chair or support.
Ly	Lying	Torso on the floor or horizontal surface.
<i>detailed posture</i>		Full postural information
U	Upright	Torso straight.
BF	Bent forward	Torso flexion angle between 20° and 60°.
BS	Strongly bent forward	Torso flexion angle greater than 60°
OS	Shoulder level work	Elbow(s) at or above shoulder level with hand(s) at or below head level.
OH	Overhead work	Hand(s) above head level.
<i>current action</i>		Goal-oriented action
Re	Reach	Moving an arm towards a target, no object in hand.
Pi	Pick	Picking-up an object, starts when touching the object, ends when arm stops moving with respect to the body.
Pl	Place	Placing an object, similar to <i>Re</i> but with an object in hand.
Rl	Release	Bringing arm back after manipulation.
Ca	Carry	Carrying an object. Starts at the end of <i>Pi</i> , ends at the beginning of <i>Pl</i> .
Fm	Fine manipulation	Dexterous manipulation of an object.
Sc	Screw	A special case of <i>Fm</i> : rotational screwing movement of the hand.
Id	Idle	Not doing anything with hands.

**Table 1.** Taxonomy of postures and actions used for the annotation of the data. For the sake of clarity, the *detailed posture* labels presented in the table correspond to the sub-categories of the EAWS worksheet only (arms and torso configuration). A full *detailed posture* label as used in the dataset is obtained by adding the label of the corresponding *general posture* as prefix, e.g., St.U for standing upright, or Kn.OH for kneeling with hands above head level.

### 3.1 File formats and organization

The dataset includes raw sensor data, videos, annotations of the postures and actions, and a description of the participants' features. The latter are stored in a *csv* file which contains : ID, age, gender, handedness, body mass, body dimensions required to create the Xsens avatar<sup>1</sup>, additional Plug-In Gait measurements<sup>\*\*</sup>, reference of the set-up used (Fig. 2), and reference number of the 3 sequences of activities performed. Sensor data are available in *csv* format and, when existing, in proprietary format and in standard motion analysis format (*c3d* and *bvh*). Videos of both cameras and videos of the Xsens avatar used for the annotations are available in *mp4* format. Note that the start and end of the Xsens videos are exactly aligned with the start and end of

Data	Format
Participants' features	<i>csv</i>
Xsens	<i>mvn; mvnx; c3d; bvh; csv</i>
Qualisys	<i>qtm; c3d; csv</i>
E-glove	<i>csv</i>
Videos	<i>mp4</i>
Annotations	<i>csv</i>

**Table 2.** File formats available for each type of data.

the corresponding Xsens data, whereas it is not the case for the videos of the cameras. Annotations are stored in *csv* format, with labels of the 3 annotators and the 3 levels of the taxonomy within a same file. Table 2 summarizes the file formats available for each data type. Sensor data contain the following information:

- Qualisys
  - 3D positions of the 43 markers placed on the body of participants.
- Xsens
  - Magnetic field, 3D linear acceleration, and orientation (quaternion) of the 17 sensors<sup>††</sup>.
  - 3D position, orientation (quaternion), linear and angular velocity, and linear and angular acceleration of the origin of the 23 body segments of the Xsens avatar.
  - Angles of the 22 joints (3 DoFs each) of the Xsens avatar.
  - 3D position of the Xsens avatar's center of mass.
- E-glove
  - Flexion angle of the first 3 fingers of the right hand.
  - Pressure force on the right palm and fingertip of the first 3 fingers of the right hand.

All *csv* files start with a header describing the nature of the data in each column and the corresponding units. Unix timestamps are used in all files, except in the Qualisys raw and motion analysis files which use machine specific timestamps and could not be converted (*qtm* and *c3d* format). Annotations files contain 2 timestamps: Unix timestamps that serve to align the annotations with sensor data, and Xsens video timestamps.

The dataset is structured as follows: first one folder for each sensor and data format (named *sensor.format*), then one folder per participant (named after the participant's ID), and finally one file per trial. Each file is named after the participant's ID, set-up reference, sequence reference, and trial number within the sequence (1 to 5). The suffix of the file is *.sensor.format* to be easily identifiable. For instance, an Xsens *csv* file is named *Participant\_1\_Setup\_A\_Seq\_1\_Trial\_1.xsens.csv*.

<sup>1</sup><https://tutorial.xsens.com/video/body-measurements>

<sup>\*\*</sup><https://docs.vicon.com/display/Nexus25/Take+subject+measurements+for+Plug-in+Gait>

<sup>††</sup>Those data are not raw data as an Xsens internal post-processing step is automatically applied during the export.

	<i>general posture</i>	<i>detailed posture</i>	<i>current action</i>
<b>Agreement (majority)</b>	92.0	81.5	81.6
<b>Transition (majority)</b>	6.1	4.4	12.0
<b>Disagreement with majority</b>	1.8	13.2	5.3
<b>Disagreement no majority</b>	0.1	0.9	1.1

**Table 3.** Consistency of labels across the 3 annotators, in percentage of time frames. *Agreement* means that all 3 annotators gave the same label. *Transition* corresponds to time frames where only 2 annotators agree, due to a variation of the transitioning instant between 2 labels. *Disagreement* corresponds to insertion or deletion of a label, or a difference in posture/action interpretation. The only situation where no majority exists is a *disagreement* where all 3 annotators assigned a different label (*disagreement no majority*).

## 3.2 Evaluation

**3.2.1 Label reliability:** Posture and action labels were manually assigned by human annotators. As in any human annotation, the labels are therefore prone to subjective bias. Hence 3 independent annotators labeled the data to reduce the effect of this bias. We computed Fleiss’  $\kappa$  to assess inter-rater reliability (Fleiss 1971). We excluded from the analysis the beginning and end of each trial, in which participants are immobile and annotation is straightforward (*i.e.*, we excluded the time frames at the beginning and end where all 3 annotators labeled the posture as *Standing Upright* and the action as *Idle*; this immobile period corresponds to the starting and stopping process of all sensors and takes up to a dozen of seconds). The results indicate that labels are consistent across annotators for all 3 levels of annotations:  $\kappa = 0.91$  for the *general posture* level,  $\kappa = 0.84$  for *detailed posture*, and  $\kappa = 0.85$  for *current action*. In addition, Table 3 presents the percentages of agreement and different disagreement types for the 3 levels of annotations. Importantly, total disagreement –where no majority is reached and a ground truth label cannot be decided– happens in only 1 % of the frames.

A significant portion of the frames where full agreement is not reached correspond to transitions, *i.e.*, disagreement on the exact frame in which the participant moved from one posture/action to the next. The median duration of a transition is 80 ms (2 frames) for the *general posture* annotation level (90<sup>th</sup> percentile: 440 ms), 80 ms for *detailed posture* (90<sup>th</sup> percentile: 400 ms), and 120 ms (3 frames) for *current action* (90<sup>th</sup> percentile: 320 ms). Though common in occurrences, those *transition* disagreements therefore remain limited to a few frames each.

**3.2.2 Completeness of optical motion capture data:** The positions of the Qualisys cameras were adapted for each of the 2 setups specifically, in order to track all markers as exhaustively as possible. The wide range of postures adopted by the participants as well as the objects they manipulated however made occlusions of some markers

inevitable. Table 4 presents, for each marker, the percentage of frames in which the marker is tracked, summarized across all trials. The median of the tracking value is greater or equal to 95 % of a trial duration for 35 markers out of 43. The leg markers were especially well tracked (given the complexity of the setup), with a median value above 97.9 % for all markers, and a 5th percentile value above 90 % for 11 out of 16 markers. The RBHD head marker was not visible at all in most trials of one participant, which explains its low 5th percentile value (33.9 %). Conversely, the low tracking scores of the markers placed at the front of the torso (CLAV, STRN, and to a lesser extent RASI and LASI) result from occlusions of those markers during bent postures in all participants. Eventually, the arm markers were tracked with a median above 95 %, except for the markers placed on the ulna styloid process (external wrist markers: RWRB and LWRB), which were easily occluded during object manipulation. Those occlusions are more prominent for the right wrist, possibly due to the e-glove structure partly hiding the marker in certain hand configurations (median of RWRB: 70.3 %, 5th percentile: 10.3 %). Note that in a few trials, a marker fell during the recording: in such situations, the tracking of the fallen marker stops when it detaches from the participant’s body. Overall, the redundancy of markers ensures that the participants’ whole-body posture can be reconstructed for most frames, even though some markers are not tracked. The reconstruction of the hand and forearm orientation should however be considered carefully when wrist markers are lost. The difficulty to fully track all markers even in a simplified lab setting highlights the advantage of wearable sensors, such as IMUs, for human whole-body motion capture in cluttered environments, such as factories.

## 3.3 Applications

We envision that the dataset presented in this paper will be used by researchers developing algorithms for classifying, predicting or evaluating human motions in industrial settings, as well as researchers developing collaborative robotics solutions that aim at improving the workers’ ergonomics. Currently, the dataset has already been used in our team for automatic activity recognition (Malaisé et al. 2019), and is currently used to develop algorithms for automatic ergonomics assessment. We used Xsens and e-glove data along with the postures and actions annotations to train recognition models based on Hidden Markov Models. The outcome is a module that successfully recognizes the current human activity and automatically identifies non-ergonomic movements on-line<sup>††</sup>.

The numerous demonstrations of the industry-inspired activities, performed by people with different anthropometric features, make our dataset ideal to conduct research in probabilistic movement primitives for action representation, and more generally to learn models of human actions at whole-body level.

This dataset could also be used by the computer vision community, to test algorithms for human posture and pose tracking in a cluttered environment. The dataset notably includes challenging postures, such as being strongly bent

<sup>††</sup>See video here: <https://youtu.be/CSTsm2xJjxg>

Head Markers	RFDH	RBHD	LFHD	LBHD	Torso Markers	CLAV	C7	RBAK	STRN	T10	RASI	RPSI	LASI	LPSI		
Median	90.3	98.5	83.6	97.1	Median	66.0	100.0	100.0	62.6	100.0	86.0	100.0	87.4	100.0		
5th %ile	71.6	33.9	61.5	68.7	5th %ile	44.4	90.1	99.7	49.8	99.5	70.2	99.0	66.8	99.5		
Right arm Markers	RSHO	RUPA	RELB	RFRA	RWRA	RWRB	RFIN	Right leg Markers	RTHI	RKNE	RTIB	RANK	RHEE	RTOE	RTOA	RTOB
Median	98.4	99.3	99.0	95.4	94.6	70.3	96.5	Median	100.0	99.9	100.0	100.0	100.0	98.7	97.9	100.0
5th %ile	86.7	94.2	85.3	78.6	59.3	10.3	89.5	5th %ile	98.4	92.1	95.3	96.2	91.4	87.3	76.9	97.2
Left arm Markers	LSHO	LUPA	LELB	LFRA	LWRA	LWRB	LFIN	Left leg Markers	LTHI	LKNE	LTIB	LANK	LHEE	LTOE	LTOA	LTOB
Median	98.9	100.0	100.0	97.3	96.8	85.8	98.3	Median	100.0	100.0	100.0	100.0	100.0	99.7	99.3	100.0
5th %ile	84.4	97.6	96.6	78.1	81.4	66.4	87.7	5th %ile	94.0	94.1	98.1	96.1	80.4	85.9	79.6	96.4

**Table 4.** Percentage of frames in which the Qualisys markers are tracked by the motion capture system. Median and 5th percentile are computed across all trials of all participants. Marker names follow the Plug-in Gait Marker Placement naming (Vicon 2017).

and crouched, and occlusions due to the environment. It could therefore be of interest for benchmarking algorithms for visual estimation of human posture (Sun et al. 2019), and for learning predictive visual models of human motion from the camera pixels directly, as recently done for manipulators (Xie et al. 2019). Remarkably, each time frame of the recorded activities is annotated. This enables to track human postures in the time dimension as well (*i.e.*, track on a sequence of images), and not only in fixed images as in many popular human posture datasets such as Andriluka et al. (2014).

Finally, our dataset is expected to be useful to the robotics community to simulate human activities in industrial settings. Simulating human activities and estimating the human ergonomics can serve to synthesize efficient collaborative robots' behaviors, for instance by addressing the issues of human-aware navigation (Khambhaita and Alami 2017) and of human-robot task allocation and coordination (Shah et al. 2011).

## 4 Conclusion

This paper presents the creation and organization of a publicly available human motion dataset containing industry-oriented activities and annotated with ergonomic labels. The dataset includes 5 hours of whole-body motion capture data both from a wearable inertial system (Xsens) and from an external optical system (Qualisys), video data from 2 cameras, and finger pressure force data measured with a sensorized glove. The dataset also includes annotations of the actions and postures of the participants according to the EAWS ergonomic worksheet, which makes it unique and especially relevant for studies and data-driven developments with a focus on ergonomics. Importantly, the dataset includes a diversity of participants and several repetitions of each task per participant to account for inter as well as intra-individual movement variability. All data are available in several formats, both proprietary and non-proprietary, to facilitate their reuse by various communities and for various applications. Due to the growing interest in improving ergonomics conditions at work, specifically with robotics

and wearable assistance, on one hand, and the time and effort needed to create and annotate a dataset on the other hand, we believe that our human motion dataset will be of great interest for the broader research community.

## Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 731540 (An.Dy). The equipment used to create the database was partly funded by the CPER IT2MP of Région Grand-Est, France. The authors would like to thank Lars Fritzsche from IMK Automotive, for his help with selecting relevant activities for the dataset, and Xsens Technologies and Emphasis Telematics for the loan of their equipment used in the data collection.

## Declaration of conflicting interests

The Authors declare that there is no conflict of interest.

## References

- Andriluka M, Pishchulin L, Gehler P and Schiele B (2014) 2d human pose estimation: New benchmark and state of the art analysis. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Busch B, Toussaint M and Lopes M (2018) Planning ergonomic sequences of actions in human-robot interaction. In: *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 1916–1923.
- Cao Z, Hidalgo G, Simon T, Wei SE and Sheikh Y (2018) OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. In: *arXiv preprint arXiv:1812.08008*.
- CMU (2019) CMU graphics lab motion capture database. URL <http://mocap.cs.cmu.edu/>. Accessed: 02-28-2019.
- Coupeté E, Moutarde F and Manitsaris S (2018) Multi-users online recognition of technical gestures for natural human-robot collaboration in manufacturing. *Autonomous Robots* : 1–17.
- David G (2005) Ergonomic methods for assessing exposure to risk factors for work-related musculoskeletal disorders. *Occupational medicine* 55(3): 190–199.
- El Makrini I, Merckaert K, De Winter J, Lefebber D and Vanderborght B (2019) Task allocation for improved ergonomics in



- human-robot collaborative assembly. *Interaction Studies* 20(1): 103–134.
- Fleiss JL (1971) Measuring nominal scale agreement among many raters. *Psychological bulletin* 76(5): 378.
- Gaudez C, Gilles M and Savin J (2016) Intrinsic movement variability at work. how long is the path from motor control to design engineering? *Applied Ergonomics* 53: 71–78.
- Gopinathan S, Mohammadi P and Steil JJ (2018) Improved human-robot interaction: A manipulability based approach. In: *2018 IEEE/RAS ICRA Workshop on Ergonomic Physical Human-Robot Collaboration*.
- Guerra-Filho G and Biswas A (2012) The human motion database: A cognitive and parametric sampling of human motion. *Image and Vision Computing* 30(3): 251–261.
- Hayes B and Shah JA (2017) Interpretable models for fast activity recognition and anomaly explanation during collaborative robotics tasks. In: *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 6586–6593.
- Hignett S and McAtamney L (2000) Rapid entire body assessment (reba). *Applied ergonomics* 31(2): 201–205.
- Ivaldi S, Fritzsche L, Babic J, Stulp F, Damsgaard M, Graimann B, Bellusci G and Nori F (2017) Anticipatory models of human movements and dynamics: the roadmap of the andy project. In: *Proc. International Conf. on Digital Human Models (DHM)*.
- Karhu O, Kansi P and Kuorinka I (1977) Correcting working postures in industry: a practical method for analysis. *Applied ergonomics* 8(4): 199–201.
- Khambhaita H and Alami R (2017) Assessing the social criteria for human-robot collaborative navigation: A comparison of human-aware navigation planners. In: *IEEE RO-MAN*.
- Kim W, Lee J, Peternel L, Tsagarakis N and Ajoudani A (2018a) Anticipatory robot assistance for the prevention of human static joint overloading in human–robot collaboration. *IEEE robotics and automation letters* 3(1): 68–75.
- Kim W, Lorenzini M, Kapıcıoğlu K and Ajoudani A (2018b) Ergotac: A tactile feedback interface for improving human ergonomics in workplaces. *IEEE Robotics and Automation Letters* 3(4): 4179–4186.
- Kubota A, Iqbal T, Shah JA and Riek LD (2019) Activity recognition in manufacturing: The roles of motion capture and semg+ inertial wearables in detecting fine vs. gross motion. In: *2019 IEEE/RAS International Conference on Robotics and Automation (ICRA)*. IEEE.
- Kuehne H, Jhuang H, Garrote E, Poggio T and Serre T (2011) HMDB: a large video database for human motion recognition. In: *Proceedings of the International Conference on Computer Vision (ICCV)*.
- Malaisé A, Maurice P, Colas F and Ivaldi S (2019) Activity recognition for ergonomics assessment of industrial tasks with automatic feature selection. *IEEE Robotics and Automation Letters*.
- Mandery C, Terlemez Ö, Do M, Vahrenkamp N and Asfour T (2015) The kit whole-body human motion database. In: *2015 International Conference on Advanced Robotics (ICAR)*. IEEE, pp. 329–336.
- Marin AG, Shourijeh MS, Galibarov PE, Damsgaard M, Fritzsche L and Stulp F (2018) Optimizing contextual ergonomics models in human-robot interaction. In: *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 1–9.
- McAtamney L and Corlett EN (1993) Rula: a survey method for the investigation of work-related upper limb disorders. *Applied ergonomics* 24(2): 91–99.
- Pearce M, Mutlu B, Shah J and Radwin R (2018) Optimizing makespan and ergonomics in integrating collaborative robots into manufacturing processes. *IEEE Transactions on Automation Science and Engineering* (99): 1–13.
- Perez MA and Nussbaum MA (2006) Posture and motion variability in non-repetitive manual materials handling tasks. *Human movement science* 25(3): 409–421.
- Peternel L, Fang C, Tsagarakis N and Ajoudani A (2018) Online human muscle force estimation for fatigue management in human-robot co-manipulation. In: *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 1340–1346.
- Punnett L and Wegman DH (2004) Work-related musculoskeletal disorders: the epidemiologic evidence and the debate. *Journal of electromyography and kinesiology* 14(1): 13–23.
- Ruffieux S, Lalanne D, Mugellini E and Khaled A O (2015) *Computer Vision and Image Understanding* 131: 72–87.
- Schaub K, Caragnano G, Britzke B and Bruder R (2013) The european assembly worksheet. *Theoretical Issues in Ergonomics Science* 14(6): 616–639.
- Schneider E, Copsey S and Irastorza X (2010) *Occupational Safety and Health in Figures: Work-related Musculoskeletal Disorders in the EU-Facts and Figures*. Office for Official Publications of the European Communities.
- Shah J, Wiken J, Williams B and Breazeal C (2011) Improved human-robot team performance using chaski, a human-inspired plan execution system. *HRI*.
- Sun K, Xiao B, Liu D and Wang J (2019) Deep high-resolution representation learning for human pose estimation. In: *CVPR*.
- Unhelkar VV, Lasota PA, Tyroller Q, Buhai RD, Marceau L, Deml B and Shah JA (2018) Human-aware robotic assistant for collaborative assembly: Integrating human motion prediction with planning in time. *IEEE Robotics and Automation Letters* 3(3): 2394–2401.
- US Bureau of Labor Statistics (2016) *Nonfatal Occupational Injuries and Illnesses Requiring Days Away from Work, 2015*.
- Vicon (2017) *Plug-in-Gait Reference Guide*. Vicon Motion Systems Ltd. URL <https://docs.vicon.com/display/Nexus25/Plug-in+Gait+models+and+templates>.
- Vignais N, Miezal M, Bleser G, Mura K, Gorecky D and Marin F (2013) Innovative system for real-time ergonomic feedback in industrial manufacturing. *Applied ergonomics* 44(4): 566–574.
- Xia L, Chen C and Aggarwal J (2012) View invariant human action recognition using histograms of 3d joints. In: *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*. IEEE, pp. 20–27.
- Xie A, Ebert F, Levine S and Finn C (2019) Improvisation through physical understanding: Using novel objects as tools with visual foresight. *R:SS*.
- Yan X, Li H, Li AR and Zhang H (2017) Wearable imu-based real-time motion warning system for construction workers' musculoskeletal disorders prevention. *Automation in Construction* 74: 2–11.