Motion Capture Benchmark of Real Industrial Tasks and Traditional Crafts for Human Movement Analysis

Brenda Elizabeth Olivas-Padilla, Alina Glushkova, Sotiris Manitsaris

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 31 March 2023, accepted 19 April 2023, date of publication 24 April 2023, date of current version 27 April 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3269581

RESEARCH ARTICLE

Motion Capture Benchmark of Real Industrial Tasks and Traditional Crafts for Human Movement Analysis

BRENDA ELIZABETH OLIVAS-PADILLA, ALINA GLUSHKOVA, AND SOTIRIS MANITSARIS
Centre for Robotics, Mines Paris, Université PSL, 75006 Paris, France

Corresponding author: Brenda Elizabeth Olivas-Padilla (brenda.olivas@minesparis.psl.eu)

The research leading to these results has received funding from the CARNOT Projet Fédérateur "Usine responsable" and the Horizon 2020 Research and Innovation Programme under Grant Agreement No. 820767, CoLLaboratE project, Grant No. 822336, Mingei project, Craeft, Grant No. 101094349.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by Article 13/14 of Regulation (EU) 2016/679 (General Data Protection Regulation), and performed in line with the Collaborate and Mingei projects.

ABSTRACT Human movement analysis is a key area of research in robotics, biomechanics, and data science. It encompasses tracking, posture estimation, and movement synthesis. While numerous methodologies have evolved over time, a systematic and quantitative evaluation of these approaches using verifiable ground truth data of three-dimensional human movement is still required to define the current state of the art. This paper presents seven datasets recorded using inertial-based motion capture. The datasets contain professional gestures carried out by industrial operators and skilled craftsmen performed in real conditions in-situ. The datasets were created with the intention of being used for research in human motion modeling, analysis, and generation. The protocols for data collection are described in detail, and a preliminary analysis of the collected data is provided as a benchmark. The Gesture Operational Model, a hybrid stochastic-biomechanical approach based on kinematic descriptors, is utilized to model the dynamics of the experts' movements and create mathematical representations of their motion trajectories for analyzing and quantifying their body dexterity. The models allowed accurate generation of human professional poses and an intuitive description of how body joints cooperate and change over time through the performance of the task.

INDEX TERMS Historical crafts, human motion generation, industrial tasks, inertial sensors, motion capture datasets, real scenarios.

I. INTRODUCTION

Previous studies of human motion data helped researchers better comprehend body dynamics and their stochastic behavior. Capturing raw data from human movement performed in different contexts permits quantifying and a better understanding of the motion parameters as well as the factors that impact motor performance. By analyzing this data, hidden parameters can be revealed, useful for motion evaluation in sports, rehabilitation, and arts but also in more professional and industrial contexts for ergonomic monitoring. Professional diseases linked to ergonomy, such as Muscular Skeletal Disorders (MSDs), constitute an important issue causing negative effects not only on the operators' health but also on the productivity of a factory/workshop.

In such context, numerous motion capture initiatives have been undertaken in various fields. Most of them have been made publicly available for visualization and analysis, and they include data corresponding to different activities of human everyday life covering from ample body motions to very fine facial expressions. These datasets can be categorized based on the technologies used (marker-based or marker-less motion capture, etc.), on the activities recorded (everyday activity, sports, etc.), or on the number of users (single user vs. multiuser interaction). For example, HumanEva and MoVi [1], [2] are existing datasets that

The associate editor coordinating the review of this manuscript and approving it for publication was Giuseppe Desolda.

contain video and marker-based motion capture (MoCap) data of a single person performing ordinary activities (like walking and jogging) and sports motions. General body movements were also recorded with a monocular camera in the HMDB51 [3] dataset, including user-object interaction, human-to-human interaction, and facial expressions. A multimodal human action database MHAD [4] has been published, including limb actions recorded with a camera, accelerometers, and a microphone. Another more recent initiative has been done in the CMU dataset [5], also including the multimodal signal from food preparation activities. For motions in multiperson interactions and scenarios, the UMPM benchmark was presented [6]. Also, the KIT dataset recorded human-to-human interaction activities while manipulating various objects [7].

In all the aforementioned studies, everyday activity has monopolized the interest of researchers performing motion capture. However, in the last decade, it has become more and more interesting to use motion capture and to apply data analysis methods to scenarios inspired by a professional context where human operators perform their tasks. The table below presents recent works of industry-oriented human motion data. Several examples can be found in the construction industry since it is one of the most affected by intense physical activity.

To detect excessive load-carrying tasks, Lee et al. [8] have focused on creating a dataset with a non-invasive single IMU sensor. The recorded data has served to automatically predict load-carrying weights and postures using deep learning algorithms. Fourteen subjects were recorded performing six different carrying modes. The dataset's analysis consists of modeling, classification, and predicting load-carrying weights. Another interesting dataset was recorded in the framework of the AnDy EU project [9], where various sensors were used, such as a full-body IMU suit including a glove for finger motion, a marker-based motion capture system, a finger pressure sensor, and 2 video cameras. The subjects performed industry-oriented activities inspired by car manufacturing. The data was annotated and labeled and is intended for use by researchers developing algorithms for classifying, predicting, or evaluating human movement in industrial settings. The evaluation focuses mostly on label reliability, not movement analysis itself. The VTT-Conlot dataset includes motion data inspired by the construction industry recorded with 3 IMUs, with 13 subjects [10]. The principal goal of this dataset is to be used for activity recognition and classification. Its evaluation refers to sensor location, modalities used, and features extracted. However, contrary to previous examples cited, the VTT-ConIot validated and compared its data also with real unannotated data belonging to real workers in a real construction site (the real data is not included in the VTT-Conlot dataset). The IKEA ASM dataset is a multi-view, furniture assembly video dataset that includes depth, atomic actions, object segmentation, and human pose [11]. One of the particularities of this one is that it includes unusual human poses performed while assembling furniture, but it does not include any IMU-captured data and aims mostly at solving computer vision challenges. The WGD dataset provides data recorded with a marker-based system of subjects performing assembly line working activities [12]. A kinematic evaluation of the data has been performed, showing that the dataset can be used for human ergonomics evaluations. Finally, the Workflow Recognition (WR) dataset comprises multi-camera video sequences recorded from the production line of an automobile manufacturer. The WR dataset was created mainly to test activity and workflow recognition algorithms [13].

All the aforementioned works went beyond recording everyday activities and focused on professional tasks/gestures/postures. However, there is still a need for MoCap data that include a greater diversity of movements, particularly professional gestures captured in real-world scenarios. Most of the datasets available were recorded inside a laboratory, causing approximate measures since they may lack authenticity and are not real workplace scenarios. Thus, this paper presents datasets created to capture and study operators' and artisans' gestures in their professional settings and real environment, performed under real conditions.

The recording procedures and processing methods are detailed in this paper. Additionally, it is provided a first analysis of the seven datasets using an analytical model called the Gesture Operational Model (GOM), which was proposed in a previous work [14]. In this analysis are created interpretable motion representations based on GOM that can be used to artificially generate human movements and explain the inter-collaboration of joints during the performance of the modeled movements. The results comprise the forecasting performance measures on every dataset and a dexterity analysis of professional tasks. The dexterity analysis applies GOM's mathematical representations to describe the performance of professional gestures. Dexterity can be defined as the skill to perform a given movement or task using the hands or other body parts. In addition, a method for identifying the most significant joint motion descriptors for modeling and recognizing a set of human movements is described. This knowledge can then be utilized to determine the ideal sensor configuration for human motion recognition problems.

II. DATA ACQUISITION

This section begins with a description of the MoCap system used for recording, followed by information on the subjects and gestures captured for each dataset.

A. MOTION CAPTURE TECHNOLOGY

The BioMed bundle motion capture system from Nansense Inc.[1] was utilized to capture the gestures of industrial operators and craftsmen. The system is composed of a full-body suit with 52 IMUs strategically positioned across the torso, limbs, and hands. At a rate of 90 frames per second, the sensors measure the orientation and acceleration of body segments on the articulated spine chain, shoulders, arms, legs,

[1] Baranger Studios, Los Angeles, CA, USA.

TABLE 1. Datasets available and employed by the community.

Dataset	Technology used	Activity recorded	Year of publication	# of subjects	Type of Environment the data has been captured in
DeTECLoad	Single IMU	Construction workers load carrying tasks	2020	14	Controlled conditions in a laboratory
AnDy project	Marker based motion capture system 4 pressure sensors 2 RGB cameras Full body IMUs + glove	Industry oriented activities (screw high/middle/low, untie knot etc.)	2020	13	Controlled conditions in a laboratory
VV-Conlot	RGB camera 3 IMUs	Construction industry oriented (painting, vacuum cleaning, etc.)	2021	13	Controlled conditions in a laboratory
IKEA ASM	3 RGB cameras	Furniture assembling (screwing etc.)	2020	48	Controlled conditions in a laboratory
WGD	Marker based motion capture system with 8 cameras	Assembly line working gestures (hammering, screwing etc.)	2021	8	Controlled conditions in a laboratory
WR	High-quality surveillance multi-camera network	Workflow tasks in automobile manufacturing	2012	≈ 11	Production line of an automobile factory

TABLE 2. Overview of the generated datasets.

Dataset	Activity recorded	Location	# of subjects
TV assembly	Drilling, connecting components on a production line, etc.	Turkey, Arcelik factory	5
Airplane floater assebly	Hammering the rivet, placing the bucking bar, etc.	Romania, Romaero factory	2
Silk weaving	Jacquard weaving gestures with looms of different sizes	Germany, Krefeld silk museum	2
Glass blowing	Shaping the decanter, blowing through the blowpipe, etc.	France, Cerfav, glass blowing workshop	1
Mastic cultivation	Sweeping the soil, embroidering the tree, etc.	Greece, Chios island, Mastic museum fields (outdoor) and in controlled indoor conditions	2
Postures according to EAWS protocol	Bending, rotating the torso, etc.	Controlled conditions in a laboratory	10

and fingertips. After a recording, the Euler local joint angles on the X, Y, and Z axes are automatically calculated through the Nansense Studio's inverse kinematics solver and stored in a Biovision Hierarchy format (BVH). A BVH file is a text file comprised of two parts. The first part provides a hierarchical description of the skeleton, beginning with the root (hips) and proceeding to the extremities of each limb. The second part of the file contains, for each frame of the recording, the absolute position of the root of the skeleton and the angles of the joints defined in the first part of the BVH file.

B. SUBJECTS RECRUITED

For the creation of each dataset, industrial operators and skilled artisans consented to be recorded in their actual workplace while wearing the Nansense suit in accordance with the General Data Protection Regulation (GDPR) principles. Firstly, industrial operators from a television plant in Istanbul, Turkey, and an aerospace company in Bucharest, Romania, were captured as they carried out their professional tasks. Four healthy people, three men and one woman, participated in the MoCap recording session at the television plant. Their average age was 31.5±6.2 years, their height was 167.8±4.6 cm, and their average weight of 65.3±9.9 kg. Two male subjects participated in the MoCap session for the recordings in the aerospace company. They had an average age of 50±5 years, a height of 170±2 cm, and a weight of 77±1.4 kg. Ten healthy individuals consented to participate in MoCap recordings of potentially dangerous ergonomic postures in a neutral environment laboratory. The subjects consisted of three women and seven men. The average age was 28.7±4.6 years, with an average height of 172.9±9.2 cm, and the average weight was 70.5±12.9 kg. None of them sustained musculoskeletal injuries, and they all completed all trials in under one hour.

Gestures of skilled artisans performing three different crafts were recorded. The first is a master silk weaver recorded at a traditional jacquard workshop in Krefeld, Germany. The expert's height was 168 cm, and his weight was 62 kg. The second artisan is a master glassblower who was recorded in action during a glassblowing workshop. The glassblower's height was 177 cm, and his weight was 73 kg. Finally, two mastic farmers were recorded at

a mastic cultivation field in Chios, Greece. Their average age was 30.5±5.5 years, height 178.8±8.5 cm, and average weight 69.3±8.0 kg.

C. RECORDING OF THE PROFESSIONAL TASKS

The procedure followed for each recording is outlined next, as well as a description of each captured task. Before recording, a calibration procedure was done. The subject assumed different postures, such as I-pose or T-pose, and performed different movements, like walking or touching his fingertips, each for 10 seconds. In order to facilitate the later annotation and segmentation of the data, only operators and artisans were asked to explain each component of the task prior to the recording.

1) INDUSTRIAL-RELATED TASKS

The gestures performed in two industrial settings have been recorded, delivering natural movements while operators execute industrial tasks. The tasks were captured on-site during regular production by actual operators.

a: TELEVISION MANUFACTURING

Two tasks were recorded at a television manufacturing plant related to assembly and packaging. The set of gestures involved in each task is designated by the abbreviations **TVA** (assembly) and **TVP** (packaging). Fig. 1 illustrates some of the gestures recorded in television assembly and packaging.

The television assembly task consists of mounting electronic circuit boards to a television chassis and using a power tool to drive screws into the boards to secure them firmly. For this task it was defined the following gesture vocabulary:

- TVA_1: Reaching high with one hand, above shoulder level, to pick one component (circuit board) from a container.
- TVA_2: Reaching low with the other empty hand, below the knee level, to pick up the second component (wire) from a second container.
- TVA_3: Connecting the components and placing the board on the chassis to be screwed.
- TVA_4: Drilling four screws on the circuit board by holding the driller with the right hand and placing the screws with the left.

The final operation required stacking the completed, boxed televisions on wooden pallets and wrapping them in a plastic membrane for shipping (TVP). The following set of gestures were recorded for this task:

- TVP_1: Placing eight TVs on a wooden pallet (bottom level).
- TVP_2: Preparing to wrap the bottom level with a membrane.
- TVP_3: Wrapping the bottom level.
- TVP_4: Placing eight TVs on top of the bottom level (second level).
- TVP_5: Wrapping the second level with a plastic membrane.
- TVP_6: Placing eight TVs on top of the second level (third level).
- TVP_7: Wrapping the third level with a plastic membrane.
- TVP_8: Placing eight TVs on top of the third level (fourth level).
- TVP_9: Wrapping the fourth level with a plastic membrane.

Boxes are given to the operator through a conveyor belt. He places one box at a time onto the pallet using both hands. After stacking eight boxes on a single level, he grabs the plastic membrane with both hands and wraps them by going around them with it. After wrapping them properly, the operator proceeds to stack boxes on top of the previous one wrapped, repeating the process. The task is complete when there are four levels of boxes on the pallet.

All tasks associated with television assembly were recorded over the course of an eight-hour shift, with one subject recorded installing the circuit boards during the first half of the shift and another recorded drilling the circuit boards to the television chassis during the second half. Three subjects were recorded separately for the packaging tasks during one shift.

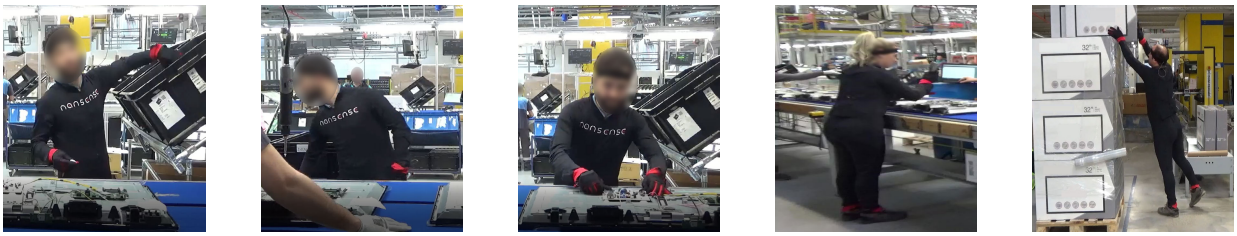
b: AIRPLANE FLOATER ASSEMBLY

The complete riveting task for an airplane floater was captured in an aerospace company. The floater is a plane component that enables planes to float when they land on water. The set of gestures recorded from this task is denoted as **APA**. Collaboration between two operators is essential for this activity. Therefore, their data were collected sequentially; one person wore the MoCap suit to capture their movement while collaborating and then donned it to the second person and continued the activity. As a result, the following gestures were recorded, which are also illustrated in Fig. 2:

- APA_1: Rivet with the pneumatic hammer.
- APA_2: Prepare the pneumatic hammer and grab rivets.
- APA_3: Place the bucking bar to counteract the incoming rivet.

One iteration of rivet assembly consisted of the first operator placing a rivet in one hole (Fig. 2a). The second operator from the opposite side of the floater then positions the bucking bar to counter the rivet (Fig. 2c). After precisely positioning the bucking bar, the second operator signals the first operator to activate the pneumatic hammer. The first operator verifies the proper placement of the assembled rivet by touching it, then moves on to the next hole and the process is repeated. After completing one line of rivets, the first operator grabs additional rivets and prepares the pneumatic hammer for the second line (Fig. 2b).

The movement of the fingers during the riveting with the pneumatic hammer was not recorded because the operator could not work realistically while wearing the MoCap gloves. The operator needed to touch with his bare hands the rivet to determine whether it was positioned correctly.



(a) TVA$_1$ (b) TVA$_2$ (c) TVA$_3$ (d) TVA$_4$ (e) TVP$_8$

FIGURE 1. Professional gestures in television manufacturing.

(a) APA$_1$ (b) APA$_2$ (c) APA$_3$

FIGURE 2. Example of airplane assembly gestures.

c: POSTURES WITH VARYING ERGONOMIC RISK LEVEL

A recording protocol was designed to capture 28 postures with varying ergonomic risk levels based on the European Assembly Worksheet (EAWS) [15].

Each posture was repeated three times, giving a total of 84 MoCap recordings per subject. The recorded postures were neutral as they were not associated with a specific activity but rather served solely to demonstrate several ergonomically incorrect postures. The postures can be divided into three main categories: those performed standing, those performed seated on a chair, and those executed while kneeling. The postures are progressing from comfortable postures to increasingly more uncomfortable but never dangerous ones. All postures were held for six seconds, and no particular discomfort was reported. This set of 28 postures with different ergonomic risk levels is denoted as **ERGD**. Three postures assumed by the subjects are shown in Fig. 3. Initially, the subject is standing with a straightened back. The subject then assumes the following three postures:

- ERGD$_1$: The subject remains standing straight up, with the arms relaxed (I-pose).
- ERGD$_2$: The subject rotates their torso to the left as far as they can for six seconds.
- ERGD$_3$: The subject bends laterally the torso to the left for six seconds.

For the next three postures, the torso is slightly bent forwards:

- ERGD$_4$: The subject remains in the bending position for six seconds.
- ERGD$_5$: While the subject is bending forward, they rotate their torso to the left and hold this position for six seconds.
- ERGD$_6$: While the subject bends forward and rotates their torso to the left, they extend their arm as if trying to reach something that is on the ground.

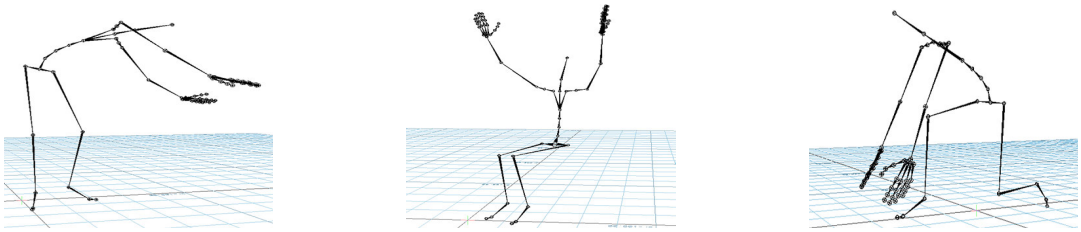
The next three postures have the torso bending forward at a large angle ($> 60°$):

- ERGD$_7$: The subject remains in the bending position for six seconds.
- ERGD$_8$: While the subject has bent forwards, they rotate their torso to the left and hold this position for six seconds.
- ERGD$_9$: While the subject bends forward and rotates their torso to the left, they extend their arm as if trying to reach something that is on the ground.

In the next few postures, the position of the arms will change, and the torso posture will be repeated:

- ERGD$_{10}$: The subject is standing upright with the forearms bend at 90° and the arms raise at the shoulder level, perpendicular to the floor.
- ERGD$_{11}$: With the arms at the same position as P10, the subject rotates their torso, and laterally bends to the left.
- ERGD$_{12}$: The participant raises their arms perpendicular to the ground while the forearms are fully extended. They proceed by rotating and laterally bending their torso to the left.
- ERGD$_{13}$: The subject raises their arms above the head for six seconds.
- ERGD$_{14}$: With the arms above the head level, the subject rotates and laterally bends to the left for six seconds.

These were all the postures that were assumed from a standing position. The next part describes the postures that will be recorded while the person is seated on a chair.



(a) ERGD$_7$ (b) ERGD$_{19}$ (c) ERGD$_{28}$

FIGURE 3. Example of postures contained in ERGD.

- ERGD$_{15}$: The person is sitting on a chair with the arms relaxed (neutral position).
- ERGD$_{16}$: While seated, the subject bends forward at an angle of 60° or more.
- ERGD$_{17}$: The subject bends forwards at an angle of 60° or more while rotating their torso and bending laterally to the left.
- ERGD$_{18}$: The subject repeats P17 but has their arms extended in front of them.
- ERGD$_{19}$: The subject raises their arms above the head level while they are fully extended.
- ERGD$_{20}$: With the arms above the head level, the participant will rotate and laterally bend their torso to the left.

Finally, the remaining postures will be performed while the subject is kneeling on their right knee. These are the most ergonomically uncomfortable postures. Beyond that, the upper body options will be the same as before:

- ERGD$_{21}$: The subject stays upright.
- ERGD$_{22}$: The subject rotates their torso to the left as far as they can, they remain in that position for six seconds.
- ERGD$_{23}$: The subject laterally bends their torso to the left.
- ERGD$_{24}$: The subject bends forward at an angle larger than 60°.
- ERGD$_{25}$: While bending the torso at an angle larger than 60°, the participant rotates and laterally bends their torso to the left.
- ERGD$_{26}$: The P25 posture is repeated, but this time, the person's arms are extended as if to pick something up from the ground.
- ERGD$_{27}$: The subject raises their arms to be perpendicular to the ground.
- ERGD$_{28}$: With the arms raised, the subject rotates and laterally bends their torso to the left.

After completing the recordings, ERGD has examples from the most comfortable positions to some of the most ergonomically improper according to the risk factors defined by EAWS. Though those postures are not in the context of any specific goal, they can act as a baseline to test different methods of an ergonomic assessment.

2) TRADITIONAL CRAFTS TASKS

Master artisans and mastic farmers were captured doing their professional tasks in their real workplaces. An additional MoCap session was conducted to capture the simulation of the mastic cultivation task without using any material or tools.

a: SILK WEAVING

In a jacquard loom workshop in Krefeld, Germany, the gestures of a skilled silk weaver were captured. This set of gestures recorded is referenced as **SLW**, and some examples of these are illustrated in Fig. 4. Throughout three days, the expert was recorded performing the following silk weaving-related tasks:

1) SLW$_1$: The creation of the punch cards.
2) SLW$_2$: Wrapping of the beam.
3) SLW$_3$: Preparation of the beam.
4) SLW$_{4,1:3}$: Jacquard weaving with looms of different sizes (small, medium, and large).

On the first day, the silk weaver was recorded performing SLW$_1$, SLW$_2$, and SLW$_3$ continuously. The creation of the punch cards was recorded for one hour. Due to the complexity and length of the tasks, the wrapping and preparation of the silk beams were recorded only once, taking about four hours to record. The next two days consisted of continuous recordings of the expert weaving using looms of three different sizes. The recording only stopped when the weaver switched to a different loom. The task of waiving with a loom can be divided into three main gestures (SLW$_{4,1}$,SLW$_{4,2}$, and SLW$_{4,3}$). Firstly, the expert pushes the pedal down with his right leg at the same time that he pushes away the threads with his left hand (the initial posture of the weaver is shown in figures 4c an 4d). Then, by controlling the shuttle that passes the thread horizontally with the right hand, he sends the shuttle to the other side with a quick pulling gesture. Finally, he pulls back the threads with the left hand while simultaneously releasing the pedal with the right leg. This process is repeated up to the end of the piece.

b: GLASSBLOWING

The creation of a glass decanter was recorded four times at Vannes-le-Châtel, France, in a European center for research and training in glasswork. Because the temperature of the glass had to be maintained throughout the process, each trial was recorded without pausing between gestures. This resulted in one motion file for each attempt, which starts with collecting the molten glass and finishes when the decanter is left to cool down. The set of gestures composing the process of creating one decanter is denoted as **GLB**.


(a) $\mathbf{SLW}_1$


(b) $\mathbf{SLW}_3$


(c) $\mathbf{SLW}_{4,3}$


(d) $\mathbf{SLW}_{4,1}$

FIGURE 4. Examples of the jacquard weaving gestures recorded.


(a) $\mathbf{GLB}_3$


(b) $\mathbf{GLB}_4$


(c) $\mathbf{GLB}_8$


(d) $\mathbf{GLB}_9$

FIGURE 5. Example of gestures captured in a glassblowing workshop.

Fig. 5 shows some of the gestures that were recorded during the decanter's fabrication. The glass decanter was created in three stages. To begin, inflate and shape the molten glass inside the decanter's main body (container). The base was created next, followed by the handle. Next, the expert rolled and shaped the decanter throughout the task to prevent the glass from deforming due to gravity. Finally, an assistant was necessary to blow into the glass while the expert shaped the decanter's main body.

For shaping the molten glass, the glassblower constantly rotated with his left hand the blowpipe while shaping the glass with his right hand. He utilized various tools with his right hand, including a block (Fig. 5b), jacks (Fig. 5c), soffietta, shears, and metal pencils. These were employed to give the glass the form of the decanter and to add further decorative details. The block is used to maintain the glass's round shape. The jacks are used to shape the decanter's cervix. The shears were utilized to cut the glass and form the decanter's peak. The soffietta forms the decanter's top. Metal pencils were then used to add the handle and extra glass details (cord around the neck) and make the foot (base) of the decanter. Manipulating the tools required constant movement of the right shoulder, right arm, and right forearm. At the same time, the glassblower was seated, rotating back and forth with the left hand the blowpipe on a metal structure. Moving the blowpipe on the metal structure required a small bending to keep the grip of the blowpipe. Placing the handle or shaping the cervix with the jacks required at times for the glassblower to stand up, but he kept moving the blowpipe with the left hand.

While forming the glass, the artisan frequently put the glass on the blowpipe into the furnace (Fig. 5d). He also continuously blew through the blowpipe while holding it horizontally at shoulder height with both arms to maintain the decanter's round shape (Fig. 5b). After finishing, it was passed to a punty to cool down.

c: MASTIC CULTIVATION

The cultivation of mastic was recorded in the span of three days in Chios, Greece. The first and second days' recordings were made outside, in front of a mastic tree. The recordings of the last day were simulated inside a room. Each task was divided into separate recordings due to the nature of the cultivation process. This resulted in separate MoCap files for each part of the process. In general, the cultivation of mastic was recorded realistically. However, specific tasks are, in reality, done days or weeks apart or take hours to be completed. As such, the expert was required to demonstrate the gestures briefly while remaining realistic. The gestures recorded from this cultivation process are denoted as **MSC**. Some gestures that were captured from the mastic farmer are shown in Fig. 6. The process begins with the preparation of the soil beneath the trees. So that dripping mastic can be easily collected, the earth surrounding the tree is cleaned and the terrain around the tree trunk is leveled. The farmer was recorded using two distinct tools to scrape the soil. The first is an antique agricultural tool (Amia) with a metal head and wooden handle, similar to a trowel. With this one, the farmer scraped the soil on his knees, holding the tool with his right hand. The second tool is a shovel, which allows the farmer to scrape the soil while standing. The farmer then swept the ground with a short broom (Fig. 6a). After preparing the soil, the farmer evenly distributed calcium carbonate ($CaCO_3$) on the ground to create a flat surface. For this task, the farmer

(a) MSC_3 (b) MSC_4 (c) MSC_5 (d) MSC_8

FIGURE 6. Example of gestures captured in the cultivation of mastic.

knelt and spread the white dust with his right hand while holding the container with his left (Fig. 6b).

The tree is then cut in order to obtain mastic. There are three different tools to do incisions in the tree. The first is a small tool with sharp points at the ends (Kenditiri), the second is another small tool called Timitiri, and the third is a small axe. The farmer was standing while using each tool, but he had to lean over to make the incisions in the tree. The tools were held with the right hand. The next step recorded was the gathering and harvesting of the mastic that had emerged from the tree's wounds. The farmer picked the fallen mastic using a small basket and tweezers (Fig. 6d), and then harvested more resin off the tree with a razor (Fig. 6c). Both gestures required the farmer to bend and manipulate the tool with his right hand.

The farmer wiped the soil to collect it on a metal mesh with a brush. In order to remove dust from the mastic, the mesh is continuously moved (or shifted). The use of two types of mesh was recorded. For all variants, the farmer knelt and moved the mesh with both hands. Finally, a third method for removing the dust from the mastic was recorded: throwing the mastic and dust while standing into the wind.

III. DATA PROCESSING AND SEGMENTATION

The processing of the MoCap consisted of two steps. To begin, a low pass filter was applied, followed by the correction of incorrect postures caused by electromagnetic interference or sensors drifting when the recording lasted too long, and calibration was required. A low-pass Butterworth filter was applied to the raw MoCap data to eliminate high-frequency noise. To avoid over-smoothing the data, the cut-off frequency was selected using the power spectrum density of the signal.

The MoCap system's sensors may drift or be influenced by magnetic disturbances from surrounding metallic objects during the recording process. As a result, occasionally erroneous joint angles were recorded during otherwise precise motion capture. The recordings were adjusted to correct this error using a 3D character animation software.[2] The software was used to adjust the unrealistic movements based on common sense and video feedback. After adjusting and removing noise from the MoCap data, it was segmented by gestures. Firstly, recordings were collected per task, with one recording representing a whole task; however, these recordings were later segmented by gestures. Fig. 7 illustrates an example of how the task of television assembly is segmented, extracting the gestures TVA_1, TVA_2, and TVA_3. All the tasks' repetitions in the seven datasets were segmented by gestures or postures for ERGD. A task may contain a single gesture that is performed numerous times, or it may contain additional gestures that are repeated throughout the task.

The segmentation of the television assembly and packaging is based on repetitions of the gestures given in Section II-C1. The repetitions segmented from the recordings are shown in Table 3. For the riveting task, the segmentation of the first gesture consisted of riveting and completing an entire line. The second gesture is to set up the pneumatic hammer for the next line of rivets. Lastly, the final gesture involved placing a bucking bar for an entire line of rivets. Table 4 illustrates the final segmentation. The recordings of postures with different ergonomic risk levels were segmented into repetitions. Given that ten subjects were recorded assuming 28 poses three times, segmentation produced 840 files containing one repetition of each pose. The tasks recorded from traditional crafts were segmented by single gestures (as there were repetitions). The resulting segmentation is displayed in tables 5, 6, and 7.

Only to facilitate the training of the models described in the next sections, the discontinuities of the Euler joint angles present in part of the MoCap files were reduced manually. These discontinuities are dramatic shifts between the values 180° and -180° in only certain local joint angles.

TABLE 3. Segmentation of the television assembly task.

Task	Gesture	Repetitions
Television Assembly	TVA_1	107
	TVA_2	107
	TVA_3	108
	TVA_4	157
Packaging	TVP_1	8
	TVP_2	2
	TVP_3	7
	TVP_4	5
	TVP_5	12
	TVP_6	7
	TVP_7	7
	TVP_8	4
	TVP_9	2

[2]MotionBuilder, Autodesk Inc., San Rafael, CA. USA.

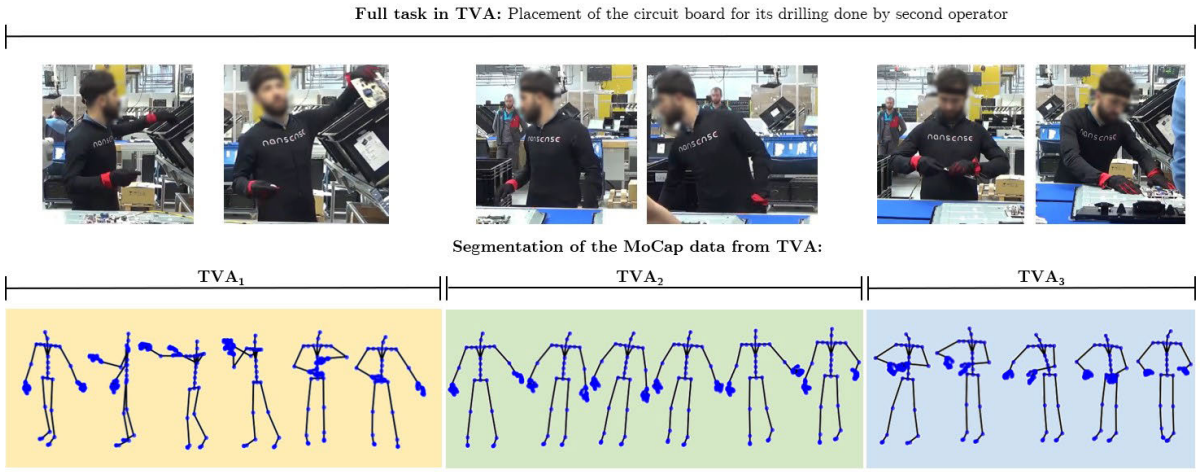


FIGURE 7. Gesture segmentation of one repetition of the task of television assembly.

TABLE 4. Segmentation of the riveting task.

Task	Gesture	Repetitions
Riveting	APA_1	6
	APA_2	5
	APA_3	8

TABLE 5. Segmentation of the silk weaving tasks.

Task	Gesture	Repetitions
Creating a card	SLW_1	110
Beam preparation	$SLW_2, 1$	3
	$SLW_2, 2$	2
	$SLW_2, 3$	4
	$SLW_2, 4$	1
	$SLW_2, 5$	1
Wrapping the beam	SLW_3	2
Weaving with small size loom	$SLW_4, 1, 1$	11
	$SLW_4, 1, 2$	11
	$SLW_4, 1, 3$	11
Weaving with medium size loom	$SLW_4, 2, 1$	35
	$SLW_4, 2, 2$	35
	$SLW_4, 2, 3$	35
Weaving with large size loom	$SLW_4, 3, 1$	16
	$SLW_4, 3, 2$	16
	$SLW_4, 3, 3$	15

TABLE 6. Segmentation of the glassblowing task.

Task	Gesture	Repetitions
Beak cutting	GLB_1	11
	GLB_2	6
Blowing and shaping	GLB_3	5
	GLB_4	8
	GLB_5	15
	GLB_6	7
	GLB_7	35
Cervix refining	GLB_8	6
Cord laying	GLB_9	2
	GLB_10	8
	GLB_11	4
Finish details	GLB_12	5
Handle laying	GLB_13	4
	GLB_14	5
	GLB_15	4
Transfer to punty	GLB_16	4
Leg and foot laying	GLB_17	6
	GLB_18	7

TABLE 7. Segmentation of the mastic cultivation task.

Task	Gesture	Repetitions
Scrapping (New tool)	MSC_1	3
Scrapping (Old tool)	MSC_2	9
Sweeping	MSC_3	9
Dusting	MSC_4	9
Embroidery A	MSC_5	9
Embroidery B	MSC_6	3
Embroidery with an axe	MSC_7	3
Gathering	MSC_8	8
Harvesting	MSC_9	7
Wiping	MSC_10	6
Shifting A	MSC_11	6
Shifting B	MSC_12	3
Cleaning with the wind	MSC_13	3

By examining each MoCap file, it was determined to transform the time series with discontinuities to a data of range $[-250°, 250°]$. Note that this transformation may not be appropriate for new movements recorded with IMUs. Nonetheless, it was sufficient to eliminate most discontinuities in the datasets presented in this paper. Each transformation was documented so that the transformed data may be inversed to Euler angles. An example of these transformations is represented in Fig. 8. The figure illustrates the MoCap data before and after the modifications, as well as the reconstructed skeleton.

The angles from the arms and forearms and one angle of the Hips were mainly the local angles with discontinuities. The angle of the Hips on the Y axis (pointing up, measuring torso rotation) was the most problematic and prone to drifting. The explanation for this could be related to the sensor's position. If the suit is loose, the sensor can produce inaccurate readings. Another factor is that after the suit is turned on and connected

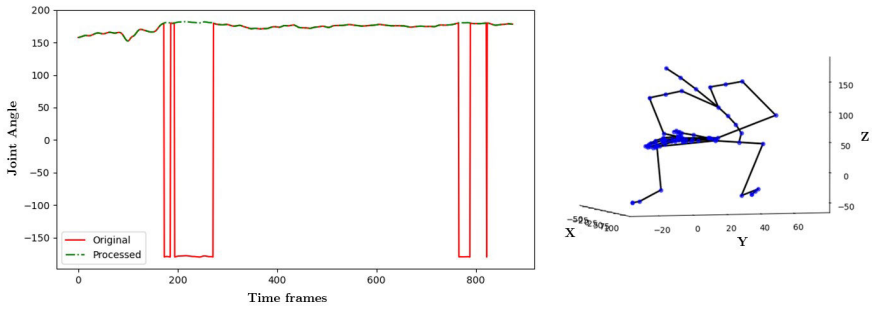


FIGURE 8. Elimination of data discontinuities for subsequent analysis. The recorded movement corresponds to the Euler angle of the left lower leg on the X-axis (shown in Fig. 9 as *LL*) for the posture $ERGD_{25}$.

to the computer for recording, the subjects must move their entire body to ''wake up'' the sensors. This sensor was most likely still in an idle state while performing calibrations. Any MoCap file with a distortion caused by drifting or poor calibration was removed from the datasets. The total size of the seven datasets utilized in the following chapters is 5GB. A total of 163,4776 frames, or 5 hours and 2 minutes, make up the segmented gestures with 156 local joint angles measured. All BVH files of the seven datasets are accessible in Zenodo.[3]

IV. ANALYSIS OF THE DATASETS USING ANALYTICAL MODELS

Any voluntary movement of the body segments is accomplished via the musculoskeletal system. The musculoskeletal system is an intricate structure comprised of bones, muscles, ligaments, and tendons. Thus, modeling a structure with such complexity is not an easy task. However, even though the musculoskeletal system is primarily responsible for the complexity of human locomotion, it can be acceptable to represent human movements using analytical models that include relevant assumptions about body joint associations and their temporal dependencies. The human movements contained in the seven datasets are then analyzed using analytical models based on the Gesture Operational Model. GOM allows quantifying human dexterity based on the learned parameters of its assumptions. The code for the analysis done in this paper is available in GitHub[4]

GOM represents human movements using a set of mathematical equations that incorporate assumptions about the stochasticity of human movement and the mediations of body joints. These assumptions allow the proper simulation of human movements using the trained models and explain the evolution of human motion descriptors across time, enabling proactive use of this information. For instance, in human-centered AI technologies, the physical embodiment of humans is the central focus (human-robot collaboration, risk monitoring, or dexterity analysis). Understanding and capturing the dependencies between the movement of different joints is crucial not only for creating more realistic human motion simulations but also for investigating how diverse and intricate full-body human movements are performed. Knowledge of the neurophysiological mechanisms behind complicated dexterity and motor learning may be gleaned from the models. Eventually, the use of such analytic models may enable the development of interdisciplinary frameworks for the research of the process of learning and skill acquisition while performing professional tasks in the industrial or craft sectors. Additionally, they might facilitate research into the key factors that lead to musculoskeletal disorders in ergonomics.

A. THE GESTURE OPERATIONAL MODEL

GOM is a mathematical representation of whole-body human movement that takes into account the spatial and temporal dynamics of body joints. The mathematical representation is comprised of a set of models, each of which models a distinct joint motion descriptor using one-shot training with Kalman filters [16]. The number of models in the equation system of GOM is equal to the number of body joints defined in GOM, multiplied by the number of dimensions the motion descriptor of each joint (e.g., angle or position) is discomposed (e.g., X, Y, and Z). For this work, the GOM was trained using motion descriptors from only 19 IMUs (out of 52 available in the datasets) for the modeling. Discarding MoCap data from the fingers and feet to simplify the human motion representation. Fig. 9 depicts the sensors' placement, labeling, and orientation. Human postures are expressed as 3D Euler joint angles in order to generate poses with subjects of various morphologies. Unlike joint positions, Euler joint angles are unaffected by identity-specific body shape. Moreover, Euler angles can be intuitively interpreted in the analytical model and provide a more clear illustration of how human movements are conducted.

Thus, 57 models compose the GOMs used in this paper to analyze the full-body movements of every dataset. The 57 models are created through state-space modeling, where endogenous and exogenous data are included in the second-order model of each motion descriptor. For example, while modeling the angle trajectory of the body joint P_t on the

[3]Benchmark website: https://doi.org/10.5281/zenodo.5356992
[4]Repository: https://github.com/olivas-bre/GOM.git

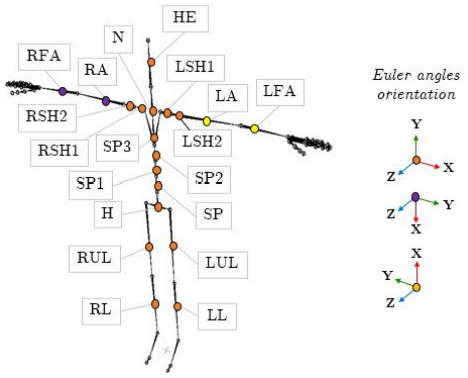


FIGURE 9. Location and Euler angle orientation of the sensors that provide the *XYZ* joint angles included in GOM.

X-axis (Px_t), whose movement is decomposed on XYZ axes (Px_t, Py_t, and Pz_t) and has an association with j body parts. The two previous values are integrated into the transition model as shown in Eq. 1, where s_t corresponds to the state variable at time t. Then, exogenous data (u_t) corresponding to potential intra-joint associations (H2), inter-limb synergies (H3), and intra-limb mediations (H4) are included in the observation model as illustrated in Eq. 2.

$$s_t = As_{t-1} = \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix} \begin{bmatrix} Px_{1,t-1} \\ -Px_{1,t-2} \end{bmatrix} \tag{1}$$

$$\begin{aligned} Px_{1,t} &= \begin{bmatrix} 1 & 1 \end{bmatrix} s_t + Bu_t \\ &= \begin{bmatrix} 1 & 1 \end{bmatrix} s_t + \beta_1 Py_{1,t-1} + \beta_2 Pz_{1,t-1} \\ &\quad + \beta_3 Px_{2,t-1} + \cdots + \beta_n Px_{j,t-1} \end{aligned} \tag{2}$$

Finally, by merging equations 1 and 2, the state-space representation of the motion descriptor is obtained:

$$\begin{aligned} Px_{1,t} = &\underbrace{\alpha_1 Px_{1,t-1} - \alpha_2 Px_{1,t-2}}_{\text{H1}} + \underbrace{\beta_1 Py_{1,t-1} + \beta_2 Pz_{1,t-1}}_{\text{H2}} \\ &+ \underbrace{\beta_3 Px_{2,t-1} + \cdots + \beta_n Px_{j,t-1}}_{\text{H3 or H4}} \end{aligned} \tag{3}$$

The assumptions and structure of the models are further detailed in [14]. The constant coefficients A and B of the equation system are estimated using Maximum Likelihood Estimation (MLE) via Kalman filtering. GOMs were trained using a reference gesture of each class, which was determined using the Dynamic Time Warping (DTW) algorithm. This algorithm measures the similarity between two time series. Then, the gesture repetition closest to all other gesture repetitions of the same class was chosen for one-shot training using an Intel Core i7-8750H CPU.

Next, Section IV-B discusses the simulation performance of the trained GOMs for every gesture in the seven datasets. Metrics and examples of poses generated are provided in the appendix. These metrics are intended to be used as an initial benchmark of the datasets for comparing the simulation or generation performance of other methods that would use the presented datasets. Later, Section IV-C presents the dexterity analysis of professional gestures using trained GOM representations and how, based on these models, the most significant motion descriptors are identified for modeling and recognizing gestures from a professional task.

B. GENERATION OF FULL-BODY MOVEMENTS

In this section are presented the results of GOM for generating human professional poses. The trained GOM can generate human professional poses by solving its equation system, with each GOM's model predicting one time step per iteration.

In order to measure the capability of the models in simulating the learned professional gestures, all gesture repetitions were simulated using their respective trained GOM. Then, the Root Mean Squared Error (RMSE) and the Mean Absolute Error (MAE) were calculated for each simulation:

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(P_t - \hat{P}_t\right)^2} \tag{4}$$

$$MAE = \frac{1}{T}\sum_{t=1}^{T}\left|P_t - \hat{P}_t\right| \tag{5}$$

The real full-body posture corresponds to P_t, and $\hat{P}_t$ is the simulated movement using the trained GOM. The average of the Theil's inequality coefficients (U_1) is also included in the metrics, which coefficient is calculated as follows:

$$U_1 = \frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(P_t - \hat{P}_t\right)^2}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}P_t^2} + \sqrt{\frac{1}{T}\sum_{t=1}^{T}\hat{P}_t^2}} \tag{6}$$

For U_1, the closer it is to zero, the greater the forecast quality. Tables 9 to 13 in the appendix show the average measures for each task's gestures. Also, figures 13 to 15 illustrate examples of generated postures of a gesture and the real posture sequence.

1) DISCUSSION ON MODEL SIMULATION

The results indicate that by solving the simultaneous equations that make up the GOM, it is possible to generate a variety of human postures using Euler joint angles as motion descriptors. GOM is tolerant of minor variations in human movement and offsets between movements of the same class resulting from varying recording conditions (different subjects or different recording days). However, suppose their performance is evaluated regarding their capability to forecast full-body movements accurately. In that case, due to the intra-class variability in some of the professional gestures, there is an increase in the mean of the joint angle errors. The reason is the potential differences between the reference gesture used for the one-shot training of the models and the testing gestures used for the simulation.

For the TVA dataset, the most difficult gestures to simulate accurately were TVA_1 and TVA_2, which corresponded to

gestures in which the operator can move more freely with either the left or right hand to grasp circuit boards or cables. On the other hand, TVA_3 and TVA_4 correspond to gestures that were easier to replicate for the operator in each iteration, as the circuit board and drilling were performed similarly in each recorded iteration.

GOM provides the best simulation performance for APA, TVP, and ERGD. These gestures and postures had the lowest intra-class variability, given that they were executed in a more controlled environment. In ERGD, for instance, subjects performed various postures in a laboratory while receiving constant instructions on how to execute them. However, as the posture became more complex, such as kneeling with torso and arm movements, the error increased as subjects' movements presented greater variation in how they performed the indicated posture (foot and knee position or arms final position). In the case of APA, the operators were recorded assembling one airplane float, performing the same tasks repeatedly for several hours with only larger variations in APA_2 where the operator grabbed the rivets and prepared the pneumatic hammer. In TVP, operators performed the same gestures with high variations only when wrapping the televisions.

The gestures recorded in industrial settings were easier to simulate since they primarily involved manipulating objects with their hands, in contrast to the gestures performed, for instance, by the craftsmen and farmers, who had to employ their entire bodies to perform their work properly.

The fact that the reference and simulated gestures were executed on different looms may have contributed to the errors in SLW (reference on a large loom and simulated on a medium-size loom). Consequently, pedal height and position variations may have caused larger errors in the movement simulation. Likewise, in the motion simulations of GLB, the skilled glassblower progressively adjusted his posture, even for the same repetitive activity, in order to appropriately shape the molten glass. Consequently, the training gestures for each class of the GLB dataset did not adequately represent all gestures from the same class (high intraclass variance), resulting in a drop in simulation accuracy. The most challenging gestures to simulate were those involved in mastic cultivation, as MSC involves gestures in which the farmer moves while kneeling. In the other six datasets, subjects performed the majority of their tasks while standing. The farmer did not keep the same position of the legs while performing the same gestures; he repositioned the legs while kneeling to improve balance in order to reach the tree or objects.

C. GOM-BASED DEXTERITY ANALYSIS OF EXPERT MOVEMENT

A statistical analysis is performed on the learned GOM representations to determine the significance of the models' assumptions in relation to the professional gesture. The significant assumptions (motion descriptors) and their learned coefficients are then used to describe the cooperation of the joints to perform the gesture. In addition, by analyzing the p-values of each assumption, the most important motion descriptors for modeling and recognizing human movements from a professional task are found. In many applications of human movement analysis, it is neither feasible nor practical to use full-body MoCap suits. Therefore, to enable the adoption of less intrusive technologies, such as smartphones and smartwatches, a procedure for finding the minimal set of motion descriptors to measure using GOM is also detailed in this paper.

1) STATISTICAL ANALYSIS AND INTERPRETATION OF THE MODELS

The statistical analysis of three trained motion representations is provided next. To facilitate the visualization of the gesture modeled, a figure with the posture sequence is provided for each example, along with color annotations to highlight the equations' assumptions. GOM's representations are designed to include four assumptions: time-dependent transitions, intra-joint association, inter-limb synergies, and serial and non-serial intra-limb mediations. Each assumption consists of a specific set of parametrized variables (in this case, joint angles) that depict a particular relationship between body joints or a temporal dependency. The notion is to use these parametrized assumptions to describe body dexterity. By examining the computed coefficients and their statistical significance (significant if the p-value is less than 0.05), it can be gleaned how relevant these are according to the gesture modeled and the predicted joint angles.

The first example illustrates the equation for the joint angle sequence RAy_t (right arm on the Y-axis) when performing the gesture TVA_1: (grab a circuit board from a container, shown in Fig. 10):

$$\begin{aligned} RAy_t = & \underbrace{(1.010)RAy_{t-1}}_{p=0.001} + \underbrace{(-0.076)RAy_{t-2}}_{p=0.188} \\ & + \underbrace{(0.720)RAx_{t-1}}_{p=0.003} + \underbrace{(1.214)RAz_{t-1}}_{p<0.001} \\ & + \underbrace{(-0.324)LAy_{t-1}}_{p<0.001} + \underbrace{(6.123)RSH1y_{t-1}}_{p<0.001} \\ & + \cdots + \underbrace{(0.555)RFAy_{t-1}}_{p=0.009} \end{aligned} \tag{7}$$

The p-values < 0.05 suggest a dependency between the prior value of the dependent variable but not between the value two time steps before. This can imply that the speed of change of the gesture is moderate. If both previous values are significant, this indicates a slow speed movement if neither is a faster one. The movement of the joint RA exhibits an intra-joint association along the X, Y, and Z axes. Inter-limb synergy with *LAy* (left arm) indicates that *LAy* follows synergistically *RAy* when performing the gesture. The movement on *RSH*1*y* (right shoulder) and *RFAy* (right forearm) result

FIGURE 10. Illustration of the gesture performed in TVA_1, where the operator grabs from a container a circuit board. The color annotations are based on the assumptions colored in (7). Colored joints indicate potential dependencies with other motion descriptors incorporated in the model: The orange indicates the transitioning assumption of *RAy*; green reflects the intra-joint association with *RAx* and *RAz*; blue highlights the inter-limb synergies with *LAy*; purple is the serial intra-limb mediations with *RSH2y* and red the non-serial intra-limb mediations with *RFAy*. The picture of the recording can also be visualized in Fig. 1a.

in a serial intra-limb mediation. This outcome makes sense, given that most of this arm movement primarily depends on shoulder motions (raising the arm). In addition, if viewing Fig. 1a, the operator must lift the shoulder and bend the forearm to reach the circuit board from the container. The bending of the forearm may explain the statistical significance of *RFAy*.

The second example is the equation for the joint angle of the neck on the X-axis (Nx_t) while performing APA_3 (hold the bucking bar, shown in Fig. 11):

$$\begin{aligned} Nx_t = &\underbrace{(1.020)\,Nx_{t-1}}_{p<0.001} + \underbrace{(0.355)\,Nx_{t-2}}_{p<0.024} \\ &+ \underbrace{(-1.220)\,Ny_{t-1}}_{p<0.001} + \underbrace{(-0.470)\,Nz_{t-1}}_{p<0.001} \\ &+ \underbrace{(-0.018)\,SP3x_{t-1}}_{p<0.001} + \underbrace{(-0.010)\,SP2x_{t-1}}_{p=0.002} \\ &+ \cdots + \underbrace{(0.010)\,Hx_{t-1}}_{p=0.84} \end{aligned} \tag{8}$$

An intra-joint association with Ny and Nz is revealed in (8), as well as a serial intra-limb mediation with SP3 (upper spine). SP2 (middle spine) exhibits non-serial intra-limb mediation, but H (hips) does not. Holding a bucking bar to counteract a rivet requires bending forward and slightly twisting the upper torso (as illustrated in figures 11 and 2c), moving along the X-axis and Y-axis of the spine. This movement is reflected in (8), as the joint angles from SP2 and SP3 on the X and Y axes are statistically significant and relevant to the motion of Nx. However, the lack of mediation with H can indicate that the operator tries to maintain his hips static, most likely to keep balance while bending. In addition, the subject had to rotate the neck to see where to position the bucking bar; thus, this is consistent with the intra-joint association indicated by the p-values of Ny and Nz. At last, the gesture is performed at a low pace as both transition assumptions are significant.

The last example is an equation learned with the gesture GLB_4 (shape the decanter curves with a block, as depicted in Fig. 12), and represents the joint angle on the X-axis of the left shoulder ($LSH2x_t$). More precisely, this equation simulates the movement of the left clavicle:

$$\begin{aligned} LSH2x_t = &\underbrace{(1.877)LSH2x_{t-1}}_{p<0.001} + \underbrace{(-0.913)LSH2x_{t-2}}_{p<0.001} \\ &+ \underbrace{(0.292)LSH2y_{t-1}}_{p=0.002} + \underbrace{(0.252)LSH2z_{t-1}}_{p=0.004} \\ &+ \underbrace{(0.145)RSH2x_{t-1}}_{p=0.014} + \underbrace{(0.36)LAx_{t-1}}_{p=0.004} \\ &+ \cdots + \underbrace{(0.016)LFAx_{t-1}}_{p=0.030} + \underbrace{(-0.543)SP3x_{t-1}}_{p=0.049} \end{aligned} \tag{9}$$

The statistical analysis of (9) reveals a temporal dependence (slow movement); intra-joint association ($LSH2y$ and $LSH2z$); inter-limb synergy with the right shoulder; serial intra-limb mediation with the left arm (LAx), and non-serial mediation with the left forearm ($LFAx$). SP3 is considered marginally significant, as this study uses a p-value threshold of 0.05 to determine significance.

To shape the decanter correctly, both arms must work together during this gesture. This is evident by the presence of an inter-limb synergy in (9). Accordingly, the joint angles of the right shoulder contribute to the response of the left shoulder, as the glassblower forms the decanter's curves with the right arm while rolling the blowpipe with the left. Furthermore, the expert mostly maintains the torso straight during this gesture, as seen in figures 13 and 5a. Yet, when he rotates the blowpipe forward, there is a slight tilt of the torso to maintain grip on the blowpipe; this could indicate a high p-value for SP3, but not as high to not be significant for the left shoulder movement.

As shown in these previous examples, GOM can provide quantitive information that is not directly observable about how the experts perform the modeled gesture and allow interpretation of how the joints collaborate to perform specific joint motion trajectories in order to perform the intended task. Calculating the significance of the assumptions highlighted

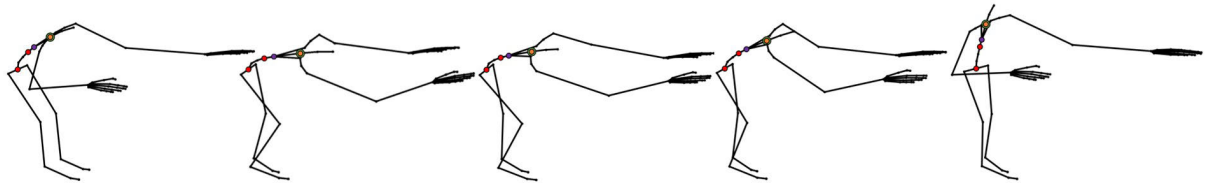


FIGURE 11. Illustration of the gesture performed in APA_3, where the operator places the bucking bar to counteract the incoming rivet. The color annotations are based on the assumptions colored in (8). Colored joints indicate potential dependencies with other motion descriptors incorporated in the model: The orange indicates the transitioning assumption of *Nx*; green reflects the intra-joint association with *Ny* and *Nz*; purple is the serial intra-limb mediations with *SP*3*x* and red the non-serial intra-limb mediations with *SP*2*x* and *Hx*. The picture of the recording can also be visualized in Fig. 2c.

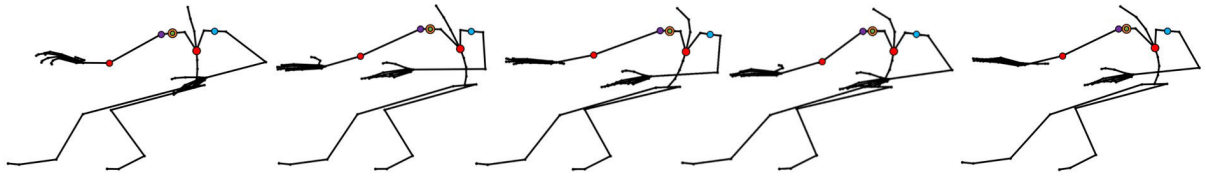


FIGURE 12. Illustration of the gesture performed in GLB_4, where the expert glassblower shapes the decanter curve with a block and simultaneously rotates the blowpipe back and forward. The color annotations are based on the assumptions colored in (9). Colored joints indicate potential dependencies with other motion descriptors incorporated in the model: The orange indicates the transitioning assumption of *LSH*2*x*; green reflects the intra-joint association with *LSH*2*y* and *LSH*2*z*; blue highlights the inter-limb synergies with *RSH*2*x*; purple is the serial intra-limb mediations with *LAx* and red the non-serial intra-limb mediations with *LFAx* and *SP*3*x*. The picture of the recording is shown in Fig. 5a.

the joints that are critical in the gesture and their influence on the movement of other joints. This information can later be utilized to test skill acquisition strategies. For example, a novice can learn to make precise gestures by minimizing the variability of their motion representations compared to those of professional artisans or operators. Moreover, for ergonomics, the proposed motion representation would allow analysts to comprehend how the full-body moves when doing ergonomically dangerous movements versus safe movements and to design work environments and tasks that are less likely to result in injury or discomfort. A direction would be to identify which joints have the highest impact while executing risky movements, and that should be monitored to reduce the ergonomic risk of the professional task.

2) SELECTION OF MOST SIGNIFICANT MOTION DESCRIPTORS PER DATASET

For selecting the essential inertial sensors to use for the gesture recognition of each professional task, the number of times a motion descriptor (assumption) is statistically significant for all equations that comprise GOM is counted. Then, different combinations of descriptors considered most frequently significant for measuring the arm, spine, and legs were utilized for training in an all-shots approach. Because a single inertial sensor gives three joint angles, all of the sensor's joint angles were used for recognition if at least one was among the joint angles that were more often significant in all gestures of a dataset.

For the recognition of human gestures utilizing different sensor combinations, Hidden Markov Models (HMM) were trained using a 10-fold cross-validation. In order to properly train the HMM, a gesture vocabulary containing the gestures with the most iterations was specified for each dataset. The total number of gesture classes for TVA, APA, and ERGD were four, three, and 28, respectively. The TVP, GLB, and MSC gesture vocabularies contained only gestures with at least seven repetitions. Therefore, their respective gesture vocabularies included five, seven, and six classes of gestures. Regarding SLW, the gesture vocabulary consisted of only three classes of silk weaving on a loom. Despite the differences in loom size, the gestures used to weave on a small, medium, and large loom are similar. Therefore, they were combined into three classes for the gesture recognition problem.

The ergodic and left-to-right HMM topologies, along with a different number of hidden states, were evaluated to determine the best settings for the gesture vocabulary defined in each dataset. The performance metrics utilized were accuracy and F1-score, the last being the harmonic mean of precision and recall. Left-to-right HMM topology produced the best results for all recognition problems. Concerning the number of hidden states, it was defined for the HMMs of TVA and ERGD with seven states, TVP with six states, APA, GLB, and MSW with eight states, and SLW with three states. The sensor configurations that obtained the best recognition results with each dataset's gestures are presented in Table 8, along with the highest recognition accuracy and F1-score achieved.

TABLE 8. Recognition performance with each configuration of sensors.

Gesture Vocabularies	# of classes	Sensors	Accuracy (%)	F1-Score (%)
TVA	4	All 52 sensors	0.967	0.966
		Two sensors: RFA, H	0.891	0.871
		LA, SP1, RUL	0.910	0.902
TVP	5	All 52 sensors	0.950	0.916
		Two sensors: RFA, H	0.888	0.850
		RSH1, LFA, SP2	0.901	0.843
APA	3	All 52 sensors	0.850	0.833
		Two sensors: RFA, H	0.750	0.701
		RA, LSH1, LSH2,SP3, SP2,LUL,RUL	0.915	0.901
ERGD	28	All 52 sensors	0.916	0.902
		Two sensors: RFA, H	0.738	0.735
		LA, RSH1, LFA,SP2, RUL	0.927	0.917
SLW	3	All 52 sensors	0.954	0.943
		Two sensors: RFA, H	0.620	0.610
		RSH1, LSH1, HE,LUL, RL	0.909	0.892
GLB	7	All 52 sensors	0.917	0.816
		Two sensors: RFA, H	0.810	0.801
		LSH2, RFA, H, SP3	0.842	0.850
MSC	6	All 52 sensors	0.866	0.866
		Two sensors: RFA, H	0.799	0.750
		LSH1, SP3, LUL, LL	0.866	0.866

In Table 8, it can be observed that using MoCap data just from the selected sensors yielded a performance that was comparable to or better than that obtained using all MoCap data from the 52 inertial sensors. In the case of TVA and TVP, just three sensors were selected based on the trained models of GOM, which represents 5.76% of the MoCap data acquired and causes only a minor loss in accuracy and F1-score. With APA and ERGD, only 13.46% and 9.61% of the MoCap data were selected and utilized to achieve a higher recognition performance than the other two configurations of sensors.

The APA gestures were the most difficult to recognize, requiring a greater number of sensors for effective recognition. This could be due to the fact that the gestures in this vocabulary are more complex and prolonged. The most problematic gesture to model and recognize was APA_2, which was expected given that its execution varied the most among the three classes (high intra-class variance). The operator did not prepare the material identically for each repetition. In certain repetitions, the operator was slower than usual because he required more time to adjust the pneumatic hammer or to prepare additional rivets. Furthermore, since only one airplane structure was built for this dataset, there is a substantial intra-class variance. There were no repetitions in which the pneumatic hammer was positioned in the same location more than once.

Regarding recognizing gestures from traditional crafts, for GLB and MSC, the selected sensors, consisting of 7.69% of the MoCap data, yielded comparable results to those obtained with data from all sensors. For SLW, 9.61% of the MoCap data was selected for the recognition problem, resulting in a performance drop of about 0.05 in both metrics with respect to the configuration with all 52 sensors. The two-sensor configuration's poor performance for SLW could be attributed to its difficulty in distinguishing movements related to the shoulder (throwing of the shuttle) and the leg, as motion data from the hips and left forearm only were insufficient.

V. CONCLUSION

This paper presented seven datasets: TVA, TVP, APA, ERGD, SLW, GLB, and MSC. Most publicly available datasets contain simulated movements performed in a laboratory and related to everyday activities or sports. Therefore, new datasets were created containing gestures performed in professional tasks either from the industry or crafts workshops. These were recorded with actual operators and experts in their real workplace scenarios using an inertial full-body suit of 52 sensors. The aim was to test human motion models with these complex gestures and extract information regarding the dexterity, skill, and know-how related to the adequate use of tangible elements such as materials and tools. Each professional task was segmented by repetitions, and discontinuities

TABLE 9. Simulation performance for datasets TVA, TVP, and APA.

Dataset	Gesture	RMSE	MAE	Avg U_1
TVA	TVA_1	52.637	25.610	0.508
	TVA_2	65.893	37.600	0.825
	TVA_3	6.836	2.742	0.116
	TVA_4	4.385	1.368	0.088
TVP	TVP_1	10.241	2.050	0.046
	TVP_2	15.746	4.779	0.078
	TVP_3	10.164	5.545	0.099
	TVP_4	9.657	2.267	0.060
	TVP_5	13.669	12.305	0.232
	TVP_6	18.192	15.816	0.202
	TVP_7	21.746	16.452	0.219
	TVP_8	16.898	4.132	0.082
	TVP_9	27.492	8.178	0.105
APA	APA_1	8.311	1.550	0.110
	APA_2	42.558	7.286	0.480
	APA_3	2.575	1.100	0.042

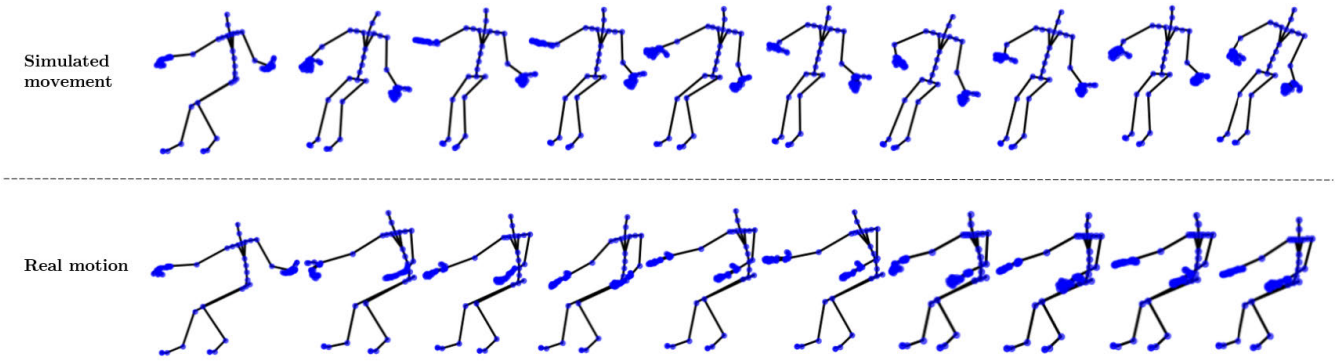


FIGURE 13. Visual comparison of generated posture sequences for GLB_4 and its ground-truth. The glassblower rotates the blowpipe with the left hand while shaping the glass with the right (the recording of the glassblower is shown in Fig. 5a).

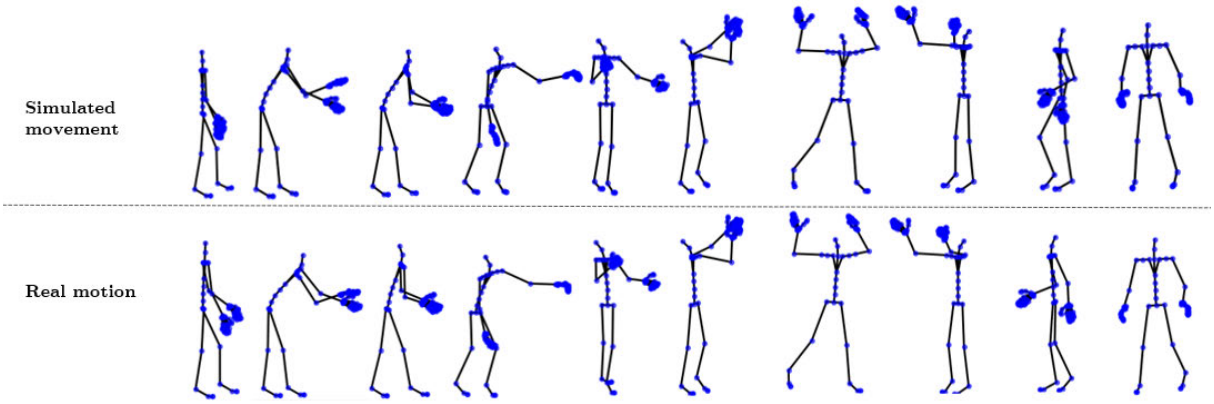


FIGURE 14. Visual comparison of generated posture sequences for TVP_8 and its ground-truth. The operator places a television on the third level of a pallet (picture of the recording in Fig. 1d).

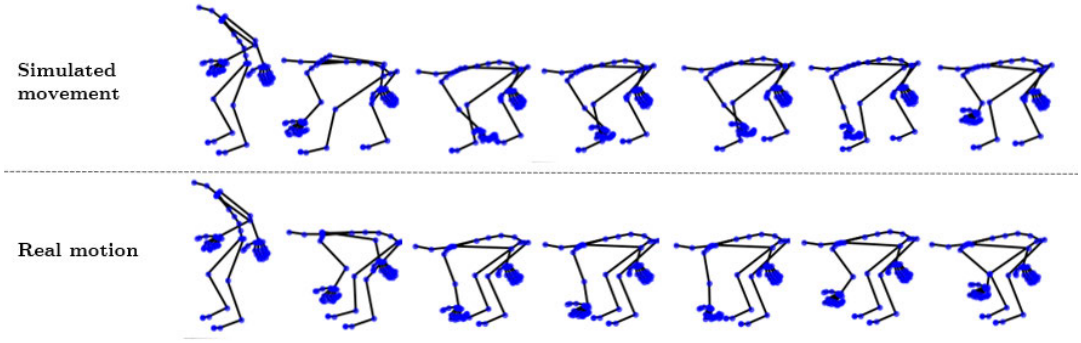


FIGURE 15. Visual comparison of generated posture sequences for MSC_5 and its ground-truth. The mastic farmer cuts the root of a mastic tree with a small knife (picture of the recording in Fig. 6c).

were reduced to improve the modeling of the gestures in the analysis done for this benchmark.

The presented human movement analysis comprised the use of GOM to simulate the recorded professional tasks and a body dexterity analysis based on the trained motion representations. The purpose was to employ the trained motion models to observe and quantify the manifestation of skill in industrial operators and expert artisans. The parameters of the train models provided information about how a person moves in order to achieve a specific goal, such as assembling a TV or making a specific piece of glass. In the future, multidisciplinary frameworks might be built to study how people learn and get better at industrial or craft tasks by looking at the trained analytical models of experts and beginners.

TABLE 10. Simulation performance for the dataset ERGD.

Dataset	Gesture	RMSE	MAE	Avg U_1
ERGD	$ERGD_1$	7.347	6.516	0.323
	$ERGD_2$	9.793	4.378	0.165
	$ERGD_3$	4.120	2.633	0.136
	$ERGD_4$	5.031	3.332	0.162
	$ERGD_5$	5.066	2.759	0.137
	$ERGD_6$	7.596	4.750	0.137
	$ERGD_7$	6.159	3.657	0.165
	$ERGD_8$	4.780	3.090	0.108
	$ERGD_9$	4.914	3.063	0.092
	$ERGD_{10}$	22.163	18.234	0.247
	$ERGD_{11}$	28.283	22.209	0.264
	$ERGD_{12}$	11.698	6.586	0.177
	$ERGD_{13}$	11.316	4.952	0.172
	$ERGD_{14}$	35.360	16.796	0.225
	$ERGD_{15}$	32.047	17.635	0.224
	$ERGD_{16}$	16.538	8.923	0.196
	$ERGD_{17}$	18.340	10.190	0.215
	$ERGD_{18}$	24.824	18.354	0.211
	$ERGD_{19}$	25.478	22.967	0.306
	$ERGD_{20}$	34.291	18.294	0.221
	$ERGD_{21}$	30.535	15.333	0.307
	$ERGD_{22}$	24.023	16.639	0.351
	$ERGD_{23}$	22.024	15.077	0.205
	$ERGD_{24}$	39.893	23.038	0.348
	$ERGD_{25}$	44.179	27.285	0.363
	$ERGD_{26}$	44.331	29.911	0.473
	$ERGD_{27}$	43.128	23.421	0.401
	$ERGD_{28}$	45.235	26.899	0.459

TABLE 11. Simulation performance for the dataset SLW.

Dataset	Gesture	RMSE	MAE	Avg U_1
SLW	SLW_1	16.518	8.841	0.063
	$SLW_{2,1}$	14.563	7.411	0.079
	$SLW_{2,2}$	14.441	7.104	0.008
	$SLW_{2,3}$	20.840	10.217	0.016
	$SLW_{2,4}$	16.551	3.097	0.005
	$SLW_{2,5}$	6.805	3.120	0.024
	SLW_3	21.907	10.666	0.082
	$SLW_{4,1,1}$	28.765	15.427	0.393
	$SLW_{4,1,2}$	27.136	13.489	0.180
	$SLW_{4,1,3}$	52.251	30.356	0.532
	$SLW_{4,2,1}$	31.695	15.635	0.234
	$SLW_{4,2,2}$	50.182	25.050	0.318
	$SLW_{4,2,3}$	47.670	18.471	0.337
	$SLW_{4,3,1}$	30.179	14.747	0.265
	$SLW_{4,3,2}$	38.361	18.740	0.293
	$SLW_{4,3,3}$	42.731	25.895	0.383

Furthermore, GOM could be used to investigate the biomechanical risk factors that lead to work-related musculoskeletal disorders by comparing motion representations from safe and hazardous movements.

Finally, the minimum number of inertial sensors and their location for capturing and accurately recognizing the gestures of each recorded professional task is presented. As stated before, employing a full-body MoCap suit in many human movement analysis applications is neither feasible nor practicable. Determining the minimal motion descriptors to measure allows for the adoption of less invasive technologies, such as smartphones and smartwatches, that could also measure these motion descriptors.

TABLE 12. Simulation performance for the dataset GLB.

Dataset	Gesture	RMSE	MAE	Avg U_1
GLB	GLB_1	68.644	36.340	0.421
	GLB_2	7.921	2.929	0.125
	GLB_3	43.928	21.410	0.294
	GLB_4	23.980	11.502	0.156
	GLB_5	49.547	24.251	0.314
	GLB_6	41.230	17.904	0.351
	GLB_7	10.292	5.213	0.146
	GLB_8	65.316	28.045	0.394
	GLB_9	62.949	21.055	0.297
	GLB_{10}	59.623	25.070	0.290
	GLB_{11}	51.223	32.240	0.534
	GLB_{12}	16.418	7.859	0.176
	GLB_{13}	38.091	18.714	0.160
	GLB_{14}	88.048	33.678	0.519
	GLB_{15}	20.144	9.203	0.198
	GLB_{16}	11.893	7.300	0.176
	GLB_{17}	46.920	15.606	0.230
	GLB_{18}	66.354	27.487	0.475

TABLE 13. Simulation performance for the dataset MSC.

Dataset	Gesture	RMSE	MAE	Avg U_1
MSC	MSC_1	4.101	2.143	0.069
	MSC_2	39.067	12.283	0.254
	MSC_3	69.123	34.171	0.468
	MSC_4	50.594	24.116	0.307
	MSC_5	52.428	25.538	0.370
	MSC_6	65.605	31.718	0.587
	MSC_7	23.665	2.197	0.030
	MSC_8	47.459	20.070	0.270
	MSC_9	55.631	26.040	0.313
	MSC_{10}	68.613	47.524	0.796
	MSC_{11}	79.191	48.210	0.841
	MSC_{12}	45.385	5.860	0.094
	MSC_{13}	68.059	23.202	0.290

APPENDIX
FORECASTING PERFORMANCE MEASURES FOR EACH DATATASET

See Tables 9–13 and Figs. 13–15.

ACKNOWLEDGMENT

We would like to thank Jean-Pierre Mateus, European Center De Recherches Et Formation Aux Arts Verriers, the Pireaus foundation, the Haus der Seidenkultur museum, and the Romaero and Arçelik factories for contributing to the creation of the datasets.

REFERENCES

[1] L. Sigal, A. O. Balan, and M. J. Black, "HumanEva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion," *Int. J. Comput. Vis.*, vol. 87, nos. 1–2, pp. 4–27, 2010.

[2] S. Ghorbani, K. Mahdaviani, A. Thaler, K. Kording, D. J. Cook, G. Blohm, and N. F. Troje, "MoVi: A large multi-purpose human motion and video dataset," *PLoS ONE*, vol. 16, no. 6, Jun. 2021, Art. no. e0253157, doi: 10.1371/journal.pone.0253157.

[3] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, and T. Serre, "HMDB: A large video database for human motion recognition," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Nov. 2011, pp. 2556–2563. [Online]. Available: http://ieeexplore.ieee.org/document/6126543/

[4] F. Ofli, R. Chaudhry, G. Kurillo, R. Vidal, and R. Bajcsy, "Berkeley MHAD: A comprehensive multimodal human action database," in *Proc. IEEE Workshop Appl. Comput. Vis.*, Clearwater, FL, USA, Jan. 2013, pp. 53–60. [Online]. Available: https://ieeexplore.ieee.org/document/6474999

[5] *CMU Graphics Lab Motion Capture Database*. Accessed: Jan. 10, 2023. [Online]. Available: http://mocap.cs.cmu.edu/
[6] N. P. van der Aa, X. Luo, G. J. Giezeman, R. T. Tan, and R. C. Veltkamp, "UMPM benchmark: A multi-person dataset with synchronized video and motion capture data for evaluation of articulated human motion and interaction," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCV Workshops)*, Nov. 2011, pp. 1264–1269.
[7] C. Mandery, O. Terlemez, M. Do, N. Vahrenkamp, and T. Asfour, "The KIT whole-body human motion database," in *Proc. Int. Conf. Adv. Robot. (ICAR)*, Jul. 2015, pp. 329–336. [Online]. Available: http://ieeexplore.ieee.org/document/7251476/
[8] H. Lee, K. Yang, N. Kim, and C. R. Ahn, "Detecting excessive load-carrying tasks using a deep learning network with a gramian angular field," *Autom. Construction*, vol. 120, Dec. 2020, Art. no. 103390. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0926580520309705
[9] P. Maurice, A. Malaisé, C. Amiot, N. Paris, G.-J. Richard, O. Rochel, and S. Ivaldi, "Human movement and ergonomics: An industry-oriented dataset for collaborative robotics," *Int. J. Robot. Res.*, vol. 38, no. 14, pp. 1529–1537, Dec. 2019. [Online]. Available: http://journals.sagepub.com/doi/10.1177/0278364919882089
[10] S.-M. Mäkela, A. Lämsä, J. S. Keränen, J. Liikka, J. Ronkainen, J. Peltola, J. Häikiö, S. Järvinen, and M. Bordallo López, "Introducing VTT-ConIot: A realistic dataset for activity recognition of construction workers using IMU devices," *Sustainability*, vol. 14, no. 1, p. 220, Dec. 2021. [Online]. Available: https://www.mdpi.com/2071-1050/14/1/220
[11] Y. Ben-Shabat, X. Yu, F. Sadat Saleh, D. Campbell, C. Rodriguez-Opazo, H. Li, and S. Gould, "The IKEA ASM dataset: Understanding people assembling furniture through actions, objects and pose," 2020, *arXiv:2007.00394*.
[12] C. Tamantini, F. Cordella, C. Lauretti, and L. Zollo, "The WGD—A dataset of assembly line working gestures for ergonomic analysis and work-related injuries prevention," *Sensors*, vol. 21, no. 22, p. 7600, Nov. 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/22/7600
[13] A. Voulodimos, D. Kosmopoulos, G. Vasileiou, E. Sardis, V. Anagnostopoulos, C. Lalos, A. Doulamis, and T. Varvarigou, "A threefold dataset for activity and workflow recognition in complex industrial environments," *IEEE Multimedia Mag.*, vol. 19, no. 3, pp. 42–52, Jul. 2012. [Online]. Available: http://ieeexplore.ieee.org/document/6272280/
[14] S. Manitsaris, G. Senteri, D. Makrygiannis, and A. Glushkova, "Human movement representation on multivariate time series for recognition of professional gestures and forecasting their trajectories," *Frontiers Robot. AI*, vol. 7, pp. 1–20, Aug. 2020.
[15] K. Schaub, G. Caragnano, B. Britzke, and R. Bruder, "The European assembly worksheet," *Theor. Issues Ergonom. Sci.*, vol. 14, no. 6, pp. 616–639, Nov. 2013.
[16] R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Eng.*, vol. 82, no. 1, pp. 35–45, 1960, doi: 10.1115/1.3662552.

BRENDA ELIZABETH OLIVAS-PADILLA received the B.S. degree in mechatronics engineering from the Monterrey Institute of Technology and Higher Education, Mexico, in 2016, the M.S. degree in electronics engineering from the National Technological Institute of Mexico, Mexico, in 2018, and the Ph.D. degree in real-time computer science, robotics, systems and control from Université PSL, France, in 2023. From 2018 to 2019, she was a Research Engineer with the Center of Robotics, Mines Paris, Université PSL, and a Ph.D. Student from 2019 to 2022. She has been a Postdoctoral Researcher with Université PSL since 2023. Her research interests include the wearable sensing and movement modeling for the human movement analysis of expert artisans and operators from manufacturing industries. In addition, she has worked on two H2020 projects, where she contributed to the capturing, processing, and the analyzing of technical movements for human learning and ergonomics.

ALINA GLUSHKOVA received the B.S. degree in information and communication from the University Paris 8, France, the dual master degree in management from the University of Paris Pantheon Assas and the University of Paris Dauphine, France, in 2012, and the Ph.D. degree in applied informatics from the University of Macedonia, Greece, in 2016. Since 2020, she has been a Researcher with the Centre for Robotics, Mines Paris, where she co-coordinates the Post-Master AIMove—"Artificial Intelligence and Movement in Robotics and Interactive Systems." Her research interests include human–computer interaction and, more specifically, on developing systems and algorithms that improve machines perception and make them capable of analyzing and evaluating human movement with the aim of providing feedback that would guide the gestural/postural performance of the learner. Her work has been applied in the analysis and correction of the ergonomics of the professional gesture, and in the learning of the expert gesture.

SOTIRIS MANITSARIS received the B.S. degree in applied mathematics from the Aristotle University of Thessaloniki, Greece, the double master's degree in local development from the University of Blaise-Pascal, France, and the Engineering School, University of Thessaly, Greece, and the Ph.D. degree in artificial intelligence from the University of Macedonia, Greece. In addition, he did three postdoctoral research in biomedical engineering, human–robot collaboration, and movement-based interactive systems. He has been the principal investigator of a number of European public and industrial projects and the Leader of the Post-Master AIMove—"Artificial Intelligence and Movement in Robotics and Interactive Systems." In 2020, he joined the International Advisory Board of the STOA Panel of the European Parliament, which puts a specific emphasis on the field of AI through its newly-established center for AI. He is currently the Deputy Director of the Centre for Robotics, Mines Paris, Université PSL. His research interests include human-centered artificial intelligence that consists of machine learning and pattern recognition concepts and methods that are applied to signals recorded from the human body and used as modalities for collaborating with intelligent machines.