



Human motion quality and accuracy measuring method for human–robot physical interactions

Tadele Belay Tuli¹ · Martin Manns¹ · Sebastian Zeller¹

Received: 22 July 2021 / Accepted: 8 June 2022 / Published online: 2 July 2022
© The Author(s) 2022

Abstract

In human–robot collaboration (HRC), human motion capture can be considered an enabler for switching autonomy between humans and robots to create efficient and safe operations. For this purpose, wearable motion tracking systems such as IMU and lighthouse-based systems have been used to transfer human joint motions into robot controller models. Due to reasons such as global positioning, drift, and occlusion, in some situations, e.g., HRC, both systems have been combined. However, it is still not clear if the motion quality (e.g., smoothness, naturalness, and spatial accuracy) is sufficient when the human operator is in the loop. This article presents a novel approach for measuring human motion quality and accuracy in HRC. The human motion capture has been implemented in a laboratory environment with a repetition of forty-cycle operations. Human motion, specifically of the wrist, is guided by the robot tool center point (TCP), which is predefined for generating circular and square motions. Compared to the robot TCP motion considered baseline, the hand wrist motion deviates up to 3 cm. The approach is valuable for understanding the quality of human motion behaviors and can be scaled up for various applications involving human and robot shared workplaces.

Keywords Motion capture · Human–robot collaboration · Human motion · Motion evaluation · Smart workplace · Jitters · Light-house system · IMU · Virtual reality

1 Introduction

In cognitive production systems, digital human modeling and motion generation are some of the core components required, e.g., for process planning. Adoption of decisions that are human-centric in production system may lead to a new paradigm according to [9]. Collaborative robots are another component of the cognitive production system requiring human understanding to efficiently complete tasks during human–robot collaboration (HRC). The ultimate goal is to enhance safety in the working space and human worker efficiency. There are different techniques for acquiring human

motions that can be semantically represented as move, reach and others using words or sentence structures (e.g., [31]). Wearable motion tracking systems such as IMUs in production environments are being evaluated to acquire motions of a human body for interfacing with robot controllers [7,23]. In a real-time motion tracking process, redundant motion measurement techniques may be required to analyze whether the system follows the desired path accurately or not. Accurate motion tracking requires a reliable motion synthesis approach that is capable of quantifying variation of motions to allow closed-loop controller designs.

Quality measures explain the smoothness and naturalness of the motion based on motion frame and explainable variance analysis, respectively. Motion quality is crucial for designing human–robot interaction strategies, specifically for smart workplaces. A smart workplace refers to a hybrid environment consisting of humans, robots, parts of products, and process descriptions. A common task in such a smart workplace includes product assembly jointly performed by a robot and human systems. In this aspect, the motion of one system can be influenced by the other. Either a robot follows a human worker, or the human worker complements the robot

✉ Tadele Belay Tuli
tadele-belay.tuli@uni-siegen.de

Martin Manns
martin.manns@uni-siegen.de

Sebastian Zeller
sekretariat-fams@mb.uni-siegen.de

¹ PROTECH - Institute of Production Technology, FAMS - Chair for Production Automation and Assembly, University of Siegen, 57076 Siegen, Germany
URL: <https://protech.mb.uni-siegen.de/fams>

motion [16,21]. A robot that can recognize a human action may collaborate to complete difficult tasks in dynamic assembly environments [35]. In most HRC applications, tasks are explicitly allocated to a robot or a human [18,28]. If we consider scenarios in which the robot adapts its motion according to human body joints, the movement's accurate replication remains challenging. The human body joints always follow varying paths during task repetitions. This makes it difficult for robots to collaborate using a predefined and static path. Unlike industrial robots, the human worker produces non-unique and less precise motions for any given repetitive process in such collaborations. In this regard, human motion modeling that adapts to spatial and temporal variations has been considered. For instance, Gaussian Mixture Modeling (GMM)[34], Motion Clustering (k-means) [13], Probabilistic Motion Modeling [1] and Deep Learning-based motion synthesis [14] are some of the common approaches.

A human following a robot's motion with physical interaction may alter the motion accuracy due to the human body applying force to the robot. In such a scenario, motion tracking is essential to explicitly identify the motion behavior of the human and the robot, particularly in assembly processes. Similarly, it can improve efficiency while empowering robots to be adaptable and intelligent. Recently, various groups have presented collaboration schemes between humans and robots. Some of these schemes can be described as a manual assembly process [15], cognitive understanding for coordinated physical interaction [6] and intentional human–robot physical collaboration [2]. It has been shown that human motion capture systems are capable of acquiring human motion behaviors which is coupled to robot behavior control. An example is gesture-based robot control. A robust method of quality measure can be advantageous for quantifying accuracy, smoothness and temporal variations. Good motion quality leads to safer or accepted human-centered workspace which is envisioned in Industry 5.0 [12].

2 Related work

Human motion capturing is used to simulate human motion behaviors in manual or collaborative tasks. For instance, human-like robots are used for replicating human motion based on one-to-one re-targeting [7]. Motion tracking systems that are generally used for human arm motion tracking can be categorized into optical, inertial, mechanical, magnetic, and acoustic techniques [10].

Optical-based motion capturing systems may use multiple cameras fixed at different locations to visualize all body parts and joints. The positions and orientations of each joint are calibrated to a fixed reference frame. Based on the motion-sensing techniques, a distinction is made between marker-less and marker-based optical systems. A marker-

less optical system such as a Kinect sensor tries to calculate the position of a person's body parts without additional aids [27,32]. On the other hand, marker-based systems use markers that are attached to a person's body. Multiple cameras are required to determine the position of the markers in the working space. Considerable amounts of cameras have been used for resolving occlusion problems. This improves motion accuracy compared to the marker-less system [10].

Inertial measurement units (IMU) are commonly considered for tracking human body motions by various groups in different applications [3,11,19]. Furthermore, some IMUs also have a magnetometer to improve the accuracy of the sensor values, although this exposes the system to changing magnetic fields. The main challenge with this technology is the retention of absolute position and drift compensation.

Different motion capturing systems (e.g., HTC Vive) have been investigated concerning their precision and accuracy. In [24], the precision and system latency of HTC Vive's position and orientation are described quantitatively. Significant changes in offset are reported when tracking data are lost in a virtual reality application. Two identical Vive systems are combined and compared with the WorldViz Precision Position Tracking (PPT) system in their work. According to [4], the HTC Vive headset and Vive trackers are used for tracking accurate and low latency human body motion for an immersive virtual reality experience. The approach used to measure the latency of the Vive tracker data is using a high-speed camera. However, it is unclear if the result is applicable in shared activities involving human and robot physical interaction.

In summary, the development of motion tracking may require the integration of various systems. Interfacing various systems may affect accuracy, reliability, controllability, or usability depending on the desired applications. From our preliminary investigation, we have considered a spatial analysis to understand how a motion capture sensor (e.g., HTC Vive) behaves when it is directly attached to the robot surface. The result exhibits significant noise (see Fig. 1), which requires advanced filtering techniques. The visualization shows before and after filtering are applied to the IMU system. The IMU system motion capture is not easy to decouple the functionality into a single sensor for measuring the independent motion of the sensor using MVN analyze. Therefore, we assume that the IMU system will exhibit the same behavior as the lighthouse system. Based on this preliminary test, we believe that placing a sensor directly on an actuating robot is susceptible to errors and noises that can arise due to system noise (e.g., drift, vibration) and occlusion. Therefore, we ruled out the sensor placement on the robot body for quality measuring. Direct mapping of sensors from the actual human to the digital human model is performed. In order to align the digital model orientation and position with the actual human, a transformation of the human joint motion has been applied. A similar approach has

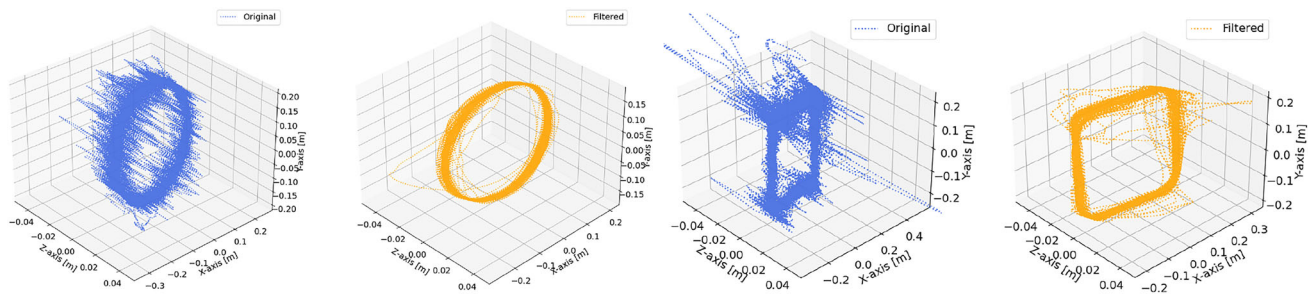


Fig. 1 Preliminary test with sensor placement on a robot TCP visualization with raw and filtered sensor data

been presented in [30,33] in which human motion tracking is implemented for running in Unity3D and robot operating system (ROS). In [36], a hidden Markov model to compensate latency of human motion has been presented. Accordingly, the authors state that the predicted motion has a root mean square error of up to 2.3 mm, considering the spatial position of the wrist joint. Here, investigating the motion capturing system's quality and accuracy, e.g., near the actuating robot, could be an interesting research question.

3 Objective

This work presents a methodology to determine the accuracy and quality of human motion capture systems for human–robot collaboration settings, in which a robot and a human jointly perform tasks and the robot takes a leading role.

The proposed application considers a joint operation of gluing a rubber strip on a car door, specifically for a window glass, which requires a circular and square motion profile. The robot is proposed to perform gluing operation, while the human worker is fixing the position following the robot's motion. Considering only the joint motions, a methodology that extends a previously presented approach, that has not considered joint activities with robots, is proposed for measuring motion capture quality [20]. In order to gain insight into its principle applicability and accuracy, it is tested with two motion capture systems and two participants of different height and size for preliminary investigation and testing. Thorough tests with a broader range of participants that may ensure general applicability is beyond the scope of this work.

The achievement of this objective could be used for:

- More realistic motion generation methods for human–robot physical interaction in situations of artifacts, occlusions, and drifts.
- Consistent motion data handling and interface approaches between human and robot models for simplifying control and motion data management.
- Digital mirrors of the physical system for developing seamless human–robot collaboration and understanding how autonomy slides between the systems.

4 Methodology

In order to develop a novel method for evaluation of motion capture accuracy and quality in HRC scenario, the following procedures are defined.

1. Define workforce requirements: For designing the experimental scenario, average workforce that fits for HRC activities is defined. Accordingly, two participants with a height between 160 and 190 cm, and Body Mass Indexes (BMIs) ranging from 20 to 30 are proposed. Age, gender, geography and so on are not in the scope of this investigation.
2. Create digital human model: For each participant, an avatar model with a skin and a kinematic model is set up.
3. Build a physical cobot setup: A cobot is set up so that it can move on a circular trajectory of 40 cm diameter and a square trajectory of 40 cm width and height. Both circle and square stand perpendicular to the floor. Their centers are positioned 160 cm above the floor. These two trajectories are chosen to be (i) simply reproducible and (ii) to cover a reasonable area of the ergonomically ideal working space (following [17]).
4. Create a digital twin of the cobot setup: Both cobot and human models are set up in one 3D environment. The human avatar is positioned so that the hip joint is situated at a 40 cm horizontal distance to the trajectory centers opposite of the robot (c. Fig. 2). From the standard idle poses of the avatars, feet positions are derived and marked on the physical shop floor for each participant.
5. Model connection between the human hand and cobot TCP: When being guided by a robot, there is a non-negligible deviation between the robot TCP and the human wrist trajectory. In preliminary tests, we have investigated this deviation for both trajectories and participants, finding systematic spatial deviations along the vertical axis in a range of 4 to 7 cm. Since this is considered too high for meaningful motion analysis, a wrist constraint is modeled that comprises an offset from the target position at the cobot. The constraint offset matches the

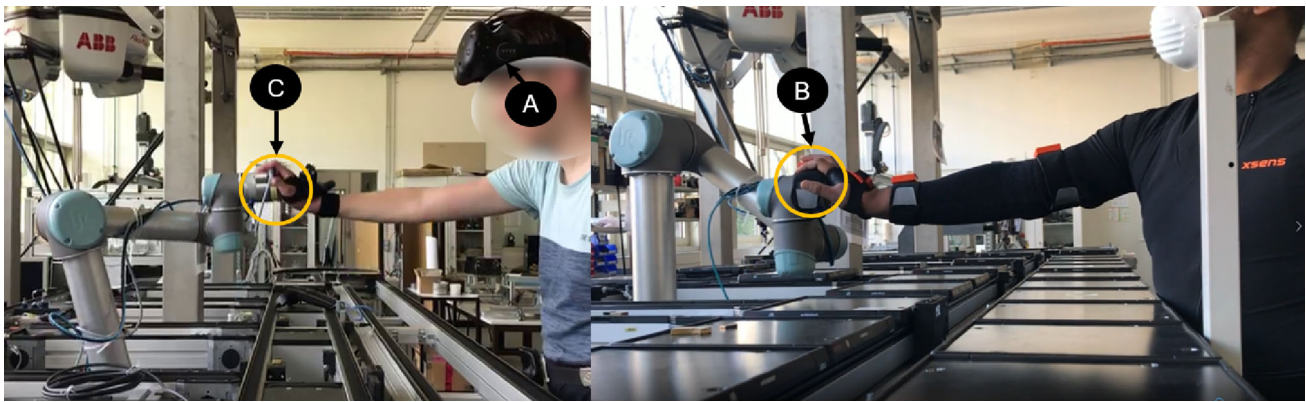


Fig. 2 Experimental setup for IMU (e.g., B) and HTC Vive (e.g., A, C)-based motion capturing process during physical interaction (e.g., C, B)

distance between the central hand joint and the wrist joint. It limits wrist joint angles to the interval $[0 \text{ deg}, 10 \text{ deg}]$. Using this method, the deviation could be reduced to 2–4 cm, measured with an Xsens IMU system during the first 30 s, in which drift stays minimal. This constraint is set up for the wrist joint of each human model in the digital twin as described. This step enables the following descriptions of HRC evaluation methods.

6. Set up motion capture system: The motion capture system that shall be tested is set up and linked to the digital model so that both robot and human motions are replicated in real time. In the case of the lighthouse system, two base stations of the HTC Vive are mounted on tripods at the height of 1.80 m, facing toward the center of the workspace, which has an action area of 1.50 m \times 3.00 m. In this aspect, the human stays inside the line of sight to avoid potential occlusions.
7. Conduct motions: For each participant, each trajectory is repeated 40 times at each take. Two takes are captured for each trajectory and participant. Joint angle data are calculated in real time from the human using the motion capture system's post-processing software and targeted to the human model in the digital twin.
8. Measuring motion quality and accuracy using spatial, motion frame and statistical methods.
9. Compare wrist trajectories: Motion captured wrist trajectories from (7) are compared with simulated ones from (8) using the FPCA approach (s. Sect. 6.2) that was presented in [20] and an RMS approach (s. Sect. 6.1).

Moreover, the following sections present details from the aspects of motion capturing, modeling (Sect. 5) and evaluation (Sect. 6).

5 Motion capture and modeling

A working space shared by humans and robots requires an interaction model. A digital working space (cf.[22]) provides

a virtual shop floor for simulation and process verification. The virtual environment, which is the digital twin of the actual system, facilitates the motion capturing system by mapping the actual joint motions into the digital model. For this matter, the digital robot model (DRM), digital human model (DHM) and kinematic model (KM) are required components that have been adapted from previous works. The DRM comprises a geometric model of a robot with kinematic chains, while DHM comprises geometrical human models with kinematic trees connecting joints and links. The KM helps for mapping the actual systems (from sensors) into the digital model using forward kinematics (FK). In the current investigation, we have implemented forward kinematics for the lighthouse-based motion tracking consisting of nine trackers. In the case of the IMU-based system, the Xsens's MVN analyzes software that has been employed to capture and stream real-time motions into the digital twin environment. Using the advantage of real-time data streaming of the robot, human joints, and tracker poses into the digital twin environment (e.g., the Unity 3D (version 2020.1.17f)), the plausibility of motions is instantly monitored (see Fig. 2). For post-processing activities, a robot operating system (ROS), which is the most common platform, has been used for capturing the states of the robot using the ROSBAG service.

While placing sensors on the human body, the motion sensors provide data of the body segment at the surface which is approximated with kinematic tree. Thus, it is necessary to use a realistic human model that minimizes the body posture errors. The DHM body postures are represented using rig bones and skin mesh, which is generated using the MakeHuman¹ software. MakeHuman is an open-source tool that allows users to create a 3D model of a person. A skeleton measurement is taken for customizing the digital model. There are 53 joints in this model, but only nine trackers in the case of HTC Vive and 17 trackers in Xsens systems are attached to the human body. The motion data are stored in a consis-

¹ <http://www.makehumancommunity.org/>.

tent file format (e.g., BVH) for easing data parsing during simulation and post-processing analysis.

A predefined path that is based on the target use case for circular and square motions is proposed to guide the human–robot joint motion. However, the interaction forces exerted by a human on the robot may affect the accuracy of the motion capture. Using ROS#² asset, the actual robot motion is transferred into the DRM through WebSocket interface. The DRM is developed based on official URDF-files from the ROS-I repository.³ The first script lets the TCP move on a circular path of 200 mm radius. With the second script, the TCP moves along a square path with an edge length of 400 mm.

6 Statistical motion evaluation and comparison

The evaluation and comparison approach employs a statistical method that is based on spatial deviations, variance of principal components and temporal variations that are further discussed in Sects. 6.1 and 6.2.

6.1 Evaluation of motion artifacts

The approach presented in [24] has described motion fluctuation using the root mean square (RMS) for observations from frame i to $i + 1$. RMS helps to measure deviations between successive frames from the captured data set. This approach yields the velocity of the deviation resulting from the change in position and orientation of the tracker and robot tool center point. The larger the value of the RMS, the larger the observed motion deviation. Similarly, the RMS is used to describe the jitters of the sensing system. The magnitude of the RMS is used to describe the impact of the jitters' artifacts.

Motion capturing of the human wrist and the robot wrist is not conducted in the same reference frame. Therefore, both systems have been transformed into a common reference frame. The data are normalized to the mean and standard deviation of 0 and 1, respectively. The transformed and normalized motion data retain its shape and the original properties of the data set.

6.2 Deviation with respect to the principal components

A principal component analysis is one of the common approaches used for identifying patterns in data explaining the similarities. By employing a functional principal component analysis (FPCA), a transformation of the raw data on

² <https://github.com/siemens/ros-sharp>.

³ https://github.com/UniversalRobots/Universal_Robots_ROS_Driver.

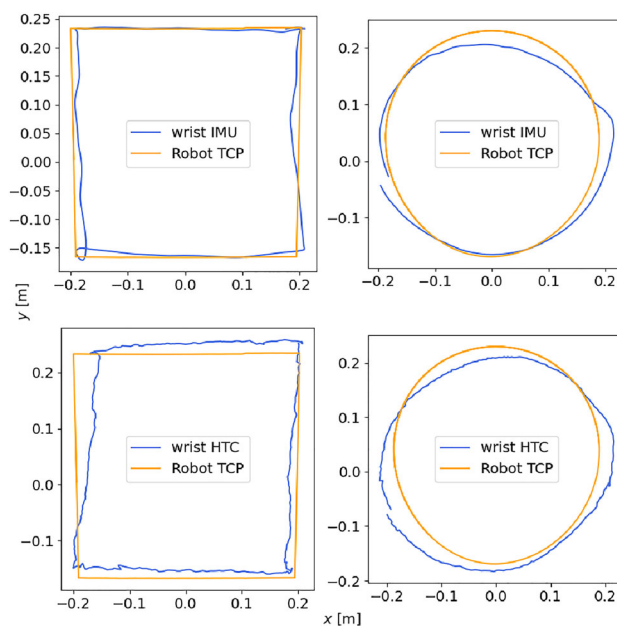


Fig. 3 The two-dimensional spatial representation and distribution of the first cycle operation

the hyperplane yields the explainable variances and eigenvalues. The explainable variances are used to measure the percentage of variation for each component. In this particular investigation, for the sake of simplicity, only joint positions are considered, which is defined by three PCA components. The eigenvalues are used to analyze how the principal components are oriented, and this is useful to compare the measured data with the reference frame. The higher the explained variation, the higher the motion naturalness (c. [8,20]). In our aspect, the motions are supposed to be planar. Therefore, any resulting deviation from a planar motion is a motion deviation that is due to unintentional human hand pressure. By employing a method developed in [25], the principal components are computed and analyzed in Sect. 7.2.2.

7 Results

The motion quality is measured based on spatial artifacts, the naturalness of the motion, and temporal deviations. Specifically, naturalness is measured based on the explainable variances of FPCA analysis, and spatial artifacts are described based on root mean square (RMS) errors. The accuracy is evaluated by comparing the captured motion with the robot trajectory.

7.1 Motion from physical interaction

A motion of two systems (i.e., human and robot) is simultaneously captured. In such cases, it is necessary to create

consistent frame rates and sampling frequencies to ensure accuracy evaluation occurs in the same space and time domain. All motions are captured at the same frame rate of 60 Hz and with the same working space and configuration. The robot is programmed to execute a square and circular motion at maximum speed. Several data sets are recorded for each experiment. The motion similarities are compared and analyzed in Sect. 7.2.

7.2 Motion comparison

The motion data visualized in Fig. 3 illustrates spatial human wrist and robot tool center point motions. It is motion-captured in real physical interaction following a defined path in collaborative mode. Single-cycle operations depicted in Fig. 3 help to visualize the distinction among all captured motions.

All these motions represent only the hand wrist joint position behavior. As is depicted in Fig. 3, the robot TCP motion is considered the reference motion or ground truth. The human wrist motion is a point on the robot's wrist surface at an offset. In the actual recording, the motion spread in all scenarios is exhibited due to the human hand posture's irregularity at each path point (see Fig. 3). The robot trajectory is straightforward for comparing the captured human motion and shows the goodness of motion qualities as described in Sect. 6.

7.2.1 Artifacts in spatial analysis

Spatial observation and representation are considered for qualitative analysis of the motion capturing systems and interaction behavior. The human hand and the robot tool center point follow a predefined path in which the robot takes the leading role. The motion from the HTC Vive in Fig. 3 has exhibited motion artifacts that are associated with jitters. The motion quality (e.g., smoothness, continuity) of the Xsens system (Fig. 3) is better than the HTC Vive and the robot motion. However, the motion is not uniform throughout the test. The robot motion relatively generates a reliable motion that is replicable.

Jitters—It is time-varying motion data that have been quantified as a displacement of peak-to-peak. It is observed in both types of motion capture systems, but it is boldly visible in the HTC Vive trackers in both test scenarios (i.e., square and circular). Jitters can be analyzed using RMS methods cf. [24]. The x-axis motion observed along the number of frames in Fig. 4 shows large signals of the motion capture (Fig. 4a, d). Similarly, the robot motion also exhibits jitters which can be due to the robot's vibration (Fig. 4c, f). By comparison, the IMU-based system shows fewer jitters (Fig. 4b, e). Compared to the preliminary investigation without a human in the loop, the quantified mean square error reduces the magnitude by 50%. The applied filtering technique

is a convolutional smoothing that applies a fixed convolution dimension on a time series using a weighted window [26].

Hand pose instability—It is difficult to maintain the human hand's position and orientation in a fixed pose during continuous and cyclic motion capturing. In addition, human body joints may occlude each other in the lighthouse-based motion capture systems. As can be observed from Fig. 3c, d, the visualization shows elastic deviation along the x-axis. This is due to the instability and inability to constrain the human hand pose. This problem is expected to occur frequently in an HRC due to disturbances or unintentional actions. Moreover, hand pose instability is a significant contributing factor to motion deviations (see Fig. 3).

Drift—The human avatar who is standing in the same place appears to be sliding across a virtual floor during a prolonged motion simulation. As a result, the motion slides along with all axes, creating an offset (see Fig. 5). The drift of IMU systems may cause this. It is difficult to decouple the drift effect from the deviation due to hand instability and the motion capturing technique. Although observation yields distinguishable behavior—e.g., the hand instability has a fluctuating pattern that is spatially constrained, however, the drift accumulates as long as the simulation takes place.

Deviation—We have measured the path length of each scenario to measure the similarities of the reference path. The circumference of the circular path of 0.2 m radius is approximately 1.23 m and the square path of 0.4 m becomes 1.60 m. Accordingly, the path length difference of the HTC Vive with human in the loop is 0.09 m beyond the planned path. The IMU system deviates up to 0.05 m. Compared to the raw sensor data without human in the loop, the path length deviates approximately 0.4 m. The filtered data have shown an improved deviation which is less than 0.1 m (see Table 1).

7.2.2 Naturalness evaluation using principal components

The third eigenvalue obtained in the FPCA computations is orthogonal to the first two, spanning the projection plane so that it equals the normal vector to the plane. FPCA for the robot is also performed to compare it with the human FPCA. The computed eigenvalues range between 0.7 and 4 degrees. And the explained variances range from 96 to 99.9% (see Table 1). The lowest explainable variance (96.61%) is obtained for circular motions of the HTC Vive system. The path length 0.8% and 7% deviation from the reference for circular motions is 2% and 4% for square motions of IMU and lighthouse systems, respectively. Compared to the preliminary investigation, the explainable variance is improved from 89 to 99.98% due to the applied filtering technique.

Fig. 4 The Violin plots for jitters observation. **a** Circular motion capture using HTC Vive, **b** circular motion using IMU system (Xsens), **c** circular motion of the robot, **d** square motion capture using HTC Vive, **e** square motion using IMU system (Xsens), **f** square motion of the robot

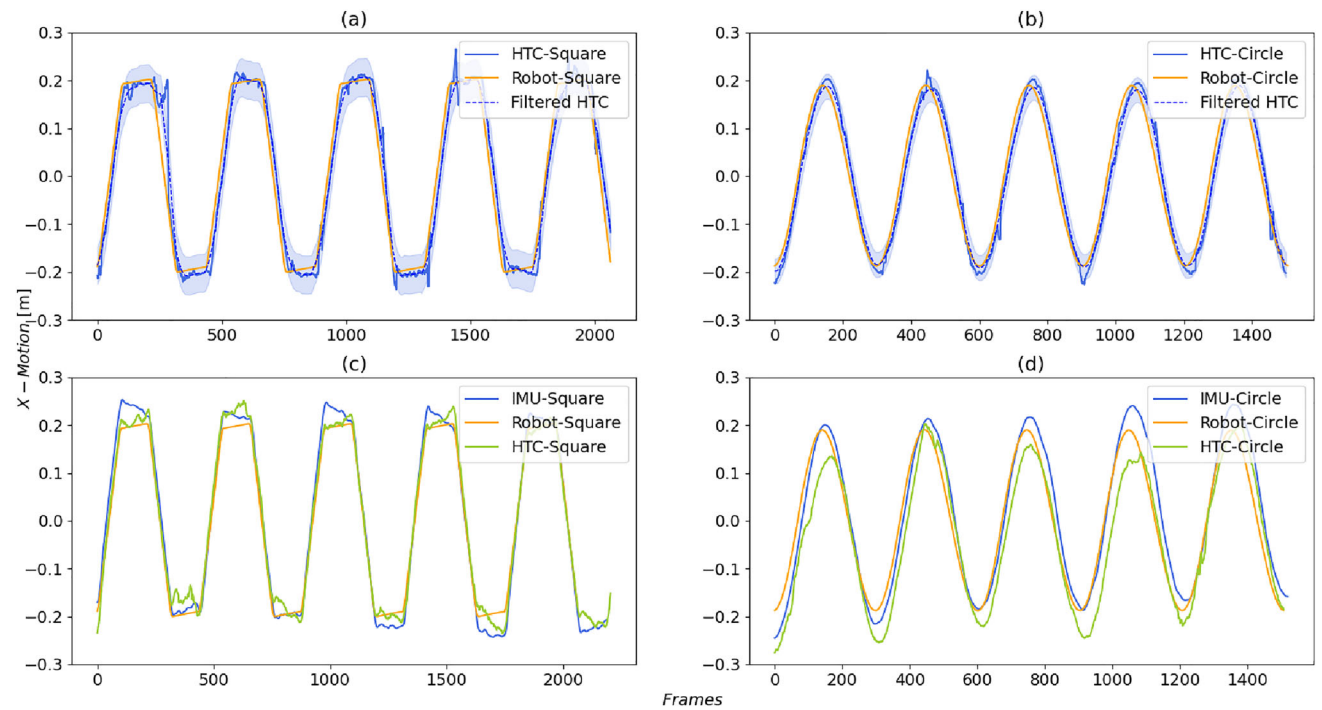
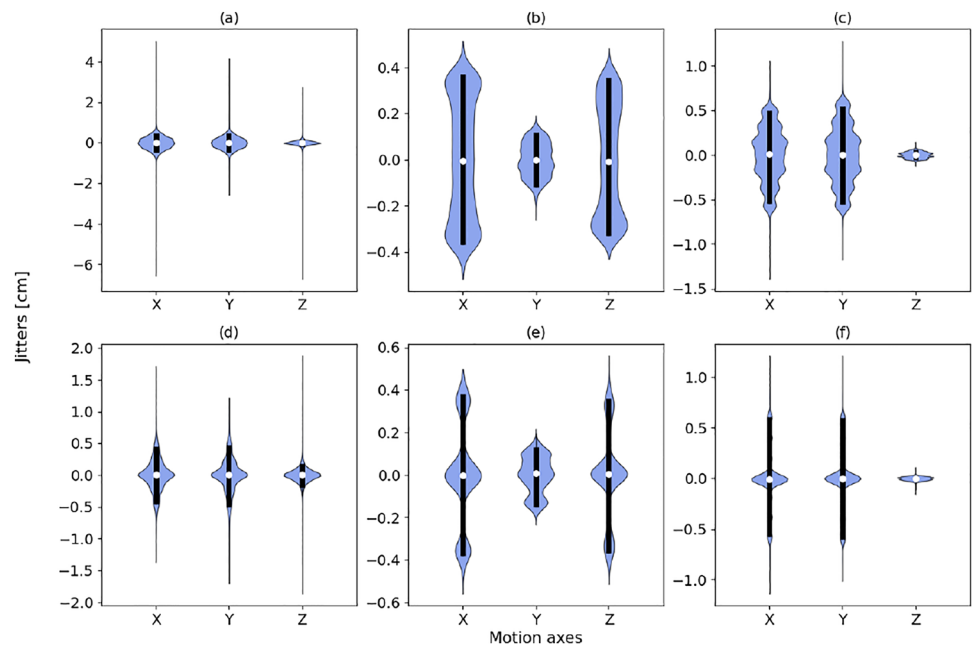


Fig. 5 Illustrative result for the first five cycles without and with human in the loop. **a** Square motion without human in the loop; **b** circular motion without human in the loop; **c** square motion with human in the loop; **d** circular motion with human in the loop

8 Discussion

According to the results described in Fig. 3, the human hand wrist does not generate the same movement pattern as the robot tool center point. It is also indicated that the robot’s motion is not unique for each cycle operation. It is also challenging to accurately position the human body joints

using wearable sensing systems because such sensors employ simplified human skeleton models and body postures. This produces significant positional offsets in the actual environment, which may affect human and robot performance during collaboration. Figure 6 illustrates the difference in joint sensing location and measured point. In this aspect, it has been required to compensate joint offset. The current investiga-

Table 1 Quality (e.g., FPCA explainable variance, RMS) and accuracy (e.g., path length deviations) measures quantitative results

Scenario	Path length [m]	Expl. var. [%]	Jitters RMS
IMU Circle	1.22	99.54	0.21
IMU Square	1.65	99.67	0.23
HTC Circle	1.32	96.61	0.26
HTC Square	1.68	99.06	0.25
HTC Square (*)	1.99	89.23	0.47
HTC Circle (*)	1.62	91.14	0.45
HTC Square (**)	1.64	99.96	0.22
HTC Circle (**)	1.32	99.98	0.20
TCP Circle	1.23	99.99	0.23
TCP Square	1.63	99.99	0.19

(*)—raw data without human in the loop, (**)—filtered data without human in the loop

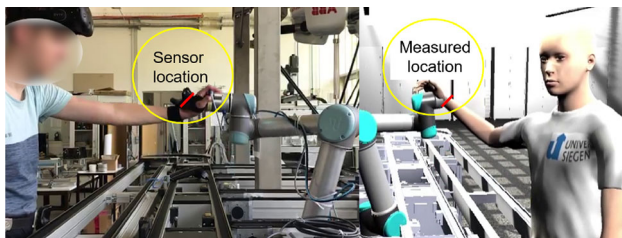


Fig. 6 Assumption in digital twin-based joint motion sensing and measuring

tion measures the motion quality and accuracy with respect to the robot tool center point (TCP). The robot's motion is more accurately captured than the human motion.

Naturalness, temporal variations and spatial artifacts are the parameters employed to measure motion qualities and accuracy of motion capture systems. Jitters and deviations measure the accuracy while FPCA measure the motion qualities. Results from jitters and deviations exhibit heterogeneous distributions for each trial. Therefore, a multi-modal distribution statistical analysis approach is employed in accordance with [5]. For a multi-modal analysis, violin plots are one of the techniques used in various works to describe the observations' distribution through graphical approaches. Accordingly, the jitters measured in the case of HTC Vive show uni-modal distribution regardless of the motion types. When we analyze the robot's jitter distribution, it is uni-modal for square motions and multi-modal distributions for circular motion (see Fig. 4). With Xsens motion capture, the circular motion is bi-modal, whereas the square is tri-modal distribution.

The potential cause for such multi-mode observation can be the Kalman filtering approaches that are implemented in MVN Analyses (c.[29]). Around turning edges (see Fig. 3), the data depict multi-peak curves that are leading to multi-mode distribution. In the case of jitters analysis (Fig. 4), the HTC Vive shows a uni-mode distribution. The Xsens-based motion curve is smoother than the robot, but the shape is

inaccurate. The smoother motion does not guarantee that the motion quality is good and accurate.

Results show that the lighthouse-based motion captures yield good positional stability with rough motions regardless of the motion type. Conversely, the IMU system generates smooth and stable motion but exhibits significant drift. The quality measure from the explainable variances shows good quality for Xsens motion data than the HTC Vive motion data. A combination of both systems may generate more robust motion as presented by Xsens.⁴ The inherent problem observed in this experiment is jitters which are dominant for HTC Vive systems and drifts in the case of the Xsens system.

The lighthouse-based motion capture is affected considerably by jitters and occlusions (see Fig. 4a, d). The motion vibration coming from the robot tool center point affects motion smoothness. By comparison, the IMU-based system generates better and smooth motion profiles. This implies that the vibration has less effect on the IMU than the lighthouse-based system. This can benefit human–robot collaborative tasks where physical interaction with a robot or the auxiliary system is desired. Hand flexibility during the operation affects both systems, but the lighthouse system shows more deviation along the normal axis.

In general, the presented approach is simple that can be easily reproduced. The approach can be scaled for an advanced motion capture setup (e.g., Vicon system cf.[11]) to measure the whole body's quality. However, it is essential to consider the equipment cost, setup time, and skill that such advanced systems require. We suggest using robot systems as a benchmark for motion capture quality measurement for a fast and economical solution. Implementation of filtering techniques such as convolutional smoothing has improved the quality of the lighthouse system, which is more or less comparable to the IMU system. However, it requires careful sensor placement, which should not be exposed to occlu-

⁴ <https://base.xsens.com/knowledgebase/s/article/Position-Aiding-HTC-Vive-1605787867640>.

sions due to workspace components or self-occlusion. The current investigation addresses jitters, deviations, and body joint flexibility (e.g., hand instability) by compensating for errors or deviations due to human and robot orientations during calibration. Similarly, a proper selection of the human location in the workplace is essential.

In future works, it will be equally important to investigate robust motion modeling methods in parallel with technological advancements to maximize motion capture systems' applicability in shop-floor environments. Such motion modeling techniques may allow robots to learn human motion behavior and predict real-time intention.

Furthermore, the proposed approach can be applied to various applications, such as automotive pre-assembly plant, gluing operation, or surface painting operations in which hand motions are desired.

9 Conclusion and outlook

Human motion capture can improve how humans and robots interact in hybrid environments. A good quality motion is crucial for establishing safe physical interactions, which may create a perception of being safe for joint operations of humans and robots. The approach to generating accurate and good motion depends on the quality of the capturing motion system and the followed procedure. Attention must be paid if direct contact between the actuating robot and tracker is considered to avoid significant jitters and drifts observed from the captured motion data. Cyclic operations with a prolonged duration are susceptible to various disturbances that can occur intentionally or unprecedentedly. The human body will mainly experience instability when attempting to maintain the pose in the same place.

In general, human motion capture in an HRC environment requires an accurate position of the human worker in the spatial space. Such an integrated system, i.e., HTC Vive and Robot using a Unity3D and ROS-I environment, may enable system controllers to enhance working space safety. However, for both tested motion capture systems, the accuracy of the motion capturing system is not better than 2 cm. Combining multiple and redundant systems such as the lighthouse and IMU-based systems can thus be regarded as a potential solution in determining working space's safety, particularly in assembly processes. In this regard, a minimal setup of low cost and easily accessible gaming tools such as HTC Vive is helpful for a virtual reality-based process demonstration and digital touring using data-driven motion capture. Future work on this topic will include investigating the simulation of HRC employing integrated IMU and the HTC Vive systems in unstructured environments.

Author Contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were per-

formed by Tadele Belay Tuli and Sebastain Zeller. The first draft of the manuscript was written by Tadele Belay Tuli and Martin Manns, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. This project is financially supported by the Federal Ministry of Education and Research of Germany within the ITEA3 project MOSIM (Grant No. 01IS18060AH) and by the European Regional Development Fund (EFRE) within the project SMAPS (Grant No. 0200545).

Availability of data and material Not applicable.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethics approval Not applicable.

Consent to participate Informed consent was obtained from all individual participants included in the study.

Consent for publication The authors affirm that human research participants provided informed consent for publication of the images in Fig. 2.

Code availability Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Agethen P, Gaisbauer F, Manns M, Link M, Rukzio E (2018) Towards realistic walk path simulation of single subjects: presenting a probabilistic motion planning algorithm. In: Proceedings of the 11th annual international conference on motion, interaction, and games—MIG '18. ACM Press, Limassol, Cyprus, pp 1–10. <https://doi.org/10.1145/3274247.3274504>. <http://dl.acm.org/citation.cfm?doid=3274247.3274504>
2. Andy Project—Home. <https://andy-project.eu/>
3. Caputo F, Greco A, D'Amato E, Notaro I, Spada S (2019) IMU-based motion capture wearable system for ergonomic assessment in industrial environment. In: Ahram TZ (ed) Advances in human factors in wearable technologies and game design, advances in intelligent systems and computing. Springer International Publishing, Berlin, pp 215–225

4. Caserman P, Garcia-Agundez A, Konrad R, Göbel S, Steinmetz R (2018) Real-time body tracking in virtual reality using a Vive tracker. *Virtual Real.* <https://doi.org/10.1007/s10055-018-0374-z>
5. Clark MW (1976) Some methods for statistical analysis of multimodal distributions and their application to grain-size data. *J Int Assoc Math Geol* 8(3):267–282. <https://doi.org/10.1007/BF01029273>
6. CoDyCo. <https://codyco.eu/>
7. Darvish K, Tirupachuri Y, Romualdi G, Rapetti L, Ferigo D, Chavez FJA, Pucci D (2019) Whole-body geometric retargeting for humanoid robots. In: 2019 IEEE-RAS 19th international conference on humanoid robots (humanoids), pp 679–686. <https://doi.org/10.1109/Humanoids43949.2019.9035059>. ISSN: 2164-0580
8. Du H, Hosseini S, Manns M, Herrmann E, Fischer K (2016) Scaled functional principal component analysis for human motion synthesis. In: Proceedings of the 9th international conference on motion in games. ACM, pp 139–144
9. ElMaraghy H, Monostori L, Schuh G, ElMaraghy W (2021) Evolution and future of manufacturing systems. *CIRP Ann* 70(2):635–658. <https://doi.org/10.1016/j.cirp.2021.05.008>
10. Field M, Pan Z, Stirling D, Naghdy F (2011) Human motion capture sensors and analysis in robotics. *Ind Robot: Int J* 38(2):163–171. <https://doi.org/10.1108/01439911111106372>
11. Filipposchi A, Schmitz N, Miezal M, Bleser G, Ruffaldi E, Stricker D (2017) Survey of motion tracking methods based on inertial sensors: a focus on upper limb human motion. *Sensors (Basel, Switzerland)* 17(6):1257. <https://doi.org/10.3390/s17061257>
12. For Research and Innovation (European Commission) DG, Breque M, De Nul L, Petridis A (2021) Industry 5.0: towards a sustainable, human centric and resilient European industry. Publications Office of the European Union, LU. <https://data.europa.eu/doi/10.2777/308407>
13. Herrmann E, Manns M, Du H, Hosseini S, Fischer K (2017) Accelerating statistical human motion synthesis using space partitioning data structures. *Comput Anim Virtual Worlds* 28(3–4):e1780. <https://doi.org/10.1002/cav.1780>
14. Holden D, Saito J, Komura T (2016) A deep learning framework for character motion synthesis and editing. *ACM Trans Graph (TOG)* 35(4):138
15. INTERACT|Interactive Manual Assembly Operations for the Human-Centered Workplaces of the Future. <http://www.interact-fp7.eu/>
16. Lenz C, Nair S, Rickert M, Knoll A, Rosel W, Gast J, Bannat A, Wallhoff F (2008) Joint-action for humans and industrial robots for assembly tasks. In: RO-MAN 2008—the 17th IEEE international symposium on robot and human interactive communication. IEEE, Munich, Germany, pp 130–135. <https://doi.org/10.1109/ROMAN.2008.4600655>. <http://ieeexplore.ieee.org/document/4600655/>
17. Lotter B (2012) überlegungen zum Montagestandort Deutschland. In: Lotter B, Wiendahl HP (eds) *Montage in der industriellen Produktion: Ein Handbuch für die Praxis*, VDI-Buch. Springer, Berlin, pp 389–396. https://doi.org/10.1007/978-3-642-29061-9_14
18. Malik AA, Bilberg A (2019) Complexity-based task allocation in human–robot collaborative assembly. *Ind Robot: Int J Robot Res Appl* 46(4):471–480. <https://doi.org/10.1108/IR-11-2018-0231>
19. Malleon C, Gilbert A, Trumble M, Collomosse J, Hilton A, Volino M (2017) Real-time full-body motion capture from video and IMUs. In: 2017 international conference on 3D vision (3DV), pp 449–457. <https://doi.org/10.1109/3DV.2017.00058>
20. Manns M, Otto M, Mauer M (2016) Measuring motion capture data quality for data driven human motion synthesis. *Procedia CIRP* 41:945–950
21. Moniz AB, Krings BJ (2016) Robots working with humans or humans working with robots? Searching for social dimensions in new human–robot interaction in industry. *Societies* 6(3):23. <https://doi.org/10.3390/soc6030023>
22. MOSIM. <https://mosim.eu/>
23. Narang S, Best A, Manocha D (2018) Simulating movement interactions between avatars & agents in virtual worlds using human motion constraints. In: 2018 IEEE conference on virtual reality and 3D user interfaces (VR). IEEE, Reutlingen, pp 9–16. <https://doi.org/10.1109/VR.2018.8446152>. <https://ieeexplore.ieee.org/document/8446152/>
24. Niehorster DC, Li L, Lappe M (2017) The accuracy and precision of position and orientation tracking in the HTC Vive virtual reality system for scientific research. *i-Perception* 8(3):23. <https://doi.org/10.1177/2041669517708205>
25. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V et al (2011) Scikit-learn: machine learning in python. *J Mach Learn Res* 12:2825–2830
26. Pelagatti MM (2021) Time series modelling with unobserved components. Taylor & Francis Limited, London (**Google-Books-ID: E6llzgEACAAJ**)
27. Peng Q, Chen W, Wu X, Wang J (2015) A novel vision-based human motion capture system using dual-Kinect. In: 2015 IEEE 10th conference on industrial electronics and applications (ICIEA), pp 51–56. <https://doi.org/10.1109/ICIEA.2015.7334083>
28. Ranz F, Hummel V, Sihn W (2017) Capability-based task allocation in human–robot collaboration. *Procedia Manuf* 9:182–189. <https://doi.org/10.1016/j.promfg.2017.04.011>
29. Roetenberg D, Luinge H, Slycke P (2009) Xsens MVN: full 6DOF human motion tracking using miniature inertial sensors, p 7
30. Su Y, Ahmadi M, Bartneck C, Steinicke F, Chen X (2019) Development of an optical tracking based teleoperation system with virtual reality. In: 2019 14th IEEE conference on industrial electronics and applications (ICIEA), pp 1606–1611. <https://doi.org/10.1109/ICIEA.2019.8833835>. ISSN: 2156-2318
31. Takano W (2020) Annotation generation from IMU-based human whole-body motions in daily life behavior. *IEEE Trans Hum-Mach Syst* 50(1):13–21. <https://doi.org/10.1109/THMS.2019.2960630>
32. Tarabini M, Marinoni M, Mascetti M, Marzaroli P, Corti F, Giberti H, Villa A, Mascagni P (2018) Monitoring the human posture in industrial environment: a feasibility study. In: 2018 IEEE sensors applications symposium (SAS), pp 1–6. <https://doi.org/10.1109/SAS.2018.8336710>
33. Tuli TB, Manns M (2019) Real-time motion tracking for humans and robots in a collaborative assembly task. *Proceedings* 42(1):48. <https://doi.org/10.3390/ecsa-6-06636> (**Number: 1 Publisher: Multidisciplinary Digital Publishing Institute**)
34. Wang JM, Fleet DJ, Hertzmann A (2008) Gaussian process dynamical models for human motion. *IEEE Trans Pattern Anal Mach Intell* 30(2):283–298
35. Wang P, Liu H, Wang L, Gao RX (2018) Deep learning-based human motion recognition for predictive context-aware human–robot collaboration. *CIRP Ann* 67(1):17–20. <https://doi.org/10.1016/j.cirp.2018.04.066>
36. Wang Q, Jiao W, Yu R, Johnson MT, Zhang Y (2019) Modeling of human welder’s operations in virtual reality human–robot interaction. *IEEE Robot Autom Lett* 4(3):2958–2964. <https://doi.org/10.1109/LRA.2019.2921928>

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.