RESEARCH ARTICLE

Learning Human-Arm Reaching Motion Using a Wearable Device in Human–Robot Collaboration

NADAV D. KAHANOWICH AND **AVISHAI SINTOV**, (Member, IEEE)

School of Mechanical Engineering, Tel Aviv University, Tel Aviv 6997801, Israel

Corresponding author: Avishai Sintov (sintov1@tauex.tau.ac.il)

ABSTRACT Many tasks performed by two humans require mutual interaction between arms such as handing-over tools and objects. In order for a robotic arm to interact with a human in the same way, it must reason about the location of the human arm in real-time. Furthermore, to acquire interaction in a timely manner, the robot must be able predict the final target of the human in order to plan and initiate motion beforehand. In this paper, we explore the use of a low-cost wearable device equipped with two inertial measurement units (IMU) for learning reaching motion in real-time applications of Human-Robot Collaboration (HRC). A wearable device can replace or be complementary to visual perception in cases of bad lighting or occlusions in a cluttered environment. We first train a neural-network model to estimate the current location of the arm. Then, we propose a novel model based on a recurrent neural-network to predict the future target of the human arm during motion in real-time. Early prediction of the target grants the robot with sufficient time to plan and initiate motion during the motion of the human. The accuracies of the models are analyzed concerning the features included in the motion representation. Through experiments and real demonstrations with a robotic arm, we show that sufficient accuracy is achieved for feasible HRC without any visual perception. Once trained, the system can be deployed in various spaces with no additional effort. The models exhibit high accuracy for various initial poses of the human arm. Moreover, the trained models are shown to provide high success rates with additional human participants not included in the training of the model.

INDEX TERMS Human–robot collaboration, intention recognition, wearable technology.

I. INTRODUCTION

When two humans perform a shared task, each has an ability to predict intentions of his peer without verbal communication. Once one human sees the motion of his human fellow, usually his arms and manipulated objects, the intended upcoming task can be predicted for further interaction [1]. For instance, when a human is handing-over an object, his fellow can infer about the reaching target and initiate supporting motion to obtain it. In *Human-Robot Collaboration* (HRC), a robot should act similarly to support a human in completing shared tasks [2]. Having a robot infer about human upper-limbs motion has applications in hand-over activities [3], [4], collaboration in shared workspaces [5], collision avoidance by the robot [6] and virtual reality [7].

Human motion inference and prediction have been given much attention in the past few decades. Seminal work by Flash and Hogen [8] has suggested that human arms move from point to point in a smooth trajectory while minimizing the mean-square jerk. Others proposed models that include minimum torque [9] and position variances [10]. Oguz et al. [11] solved an inverse optimal control problem in order to derive the true cost function that governs a set of motions. The work in [12] predicted toward which object the human hand is reaching. The target object was predicted by observing arm motion with a camera in a discretized workspace. Nevertheless, these models were demonstrated

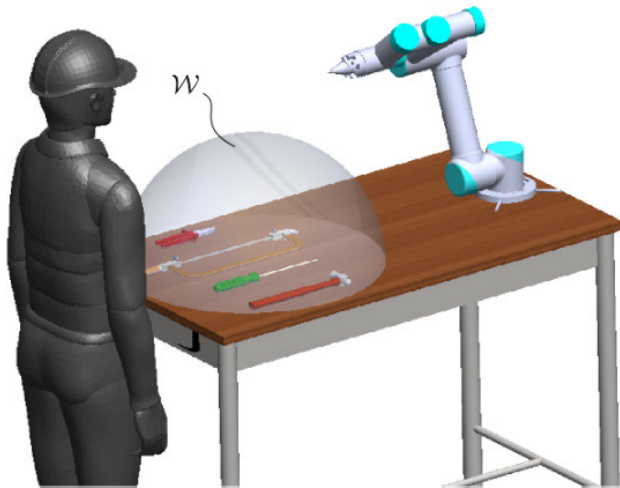The associate editor coordinating the review of this manuscript and approving it for publication was Giacinto Barresi.

FIGURE 1. A human user and a robot share space $\mathcal{W}$ where they interact or collaborate in a shared task.

in a limited and specific setting while human motion is difficult to predict due to the randomness and complexity of human behavior [13]. In addition, some of these models may require calibration to the user and the acquisition of musculoskeletal parameters that cannot be easily extracted. Therefore, a significant amount of research has been put on the use of Kalman filter variants [14], [15] and data-driven models [16], [17], [18], [19], [20].

In an example for data-driven models, the work of Landi et al. [21] combined the minimum jerk model with an Artificial Neural-Network (ANN) to predict arm motion based on camera perception. The work of Cheng et al. [19] proposed the use of semi-adaptable ANN to learn a human arm transition model and adapt it to time-varying human behaviors. Another approach used Hidden Markov Models to approximate human pose or occupied workspace based on visual observations [16], [22]. The use of Recurrent Neural-Networks (RNN) is also a common approach where sequential temporal data is used to predict motion trajectories [23], [24]. Liu and Liu [25] combined a Modified Kalman Filter to adapt an RNN model to changes in environmental conditions. The work in [17] used a database of recorded human motions in order to predict in real-time intended targets in reaching motions. It is also worth mentioning extensive attempts for pose and motion prediction of the entire human body through visual perception [26], [27], [28]. Similarly, depth camera such as Kinect is an alternative solution using spatial point cloud [29]. However, all above methods rely on visual perception in order to acquire human arm pose. Relying on continuous visual feedback limits the performance of various tasks in which visual uncertainty (e.g., poor lighting or shadows) or occlusion may occur. Moreover, dealing with visual sensing requires a large amount of data and strong computing capabilities [30]. Achieving estimation solely based on wearable sensing would enable further fusion with

vision for better estimation in unstructured or partly occluded environments.

To bypass the challenges of vision, motion prediction using wearable sensors has also been exhibited. For instance, Electro-Myography (EMG) [31] and brain-computer interface [32] have been tested for intention prediction of motor behaviour. Wearable Inertial Measurement Units (IMU) have also been proposed and yielded significant results [33]. The work in [34] fused IMU measurements with depth camera (Kinect) perception based on the Unscented Kalman Filter (UKF). Yun and Bachmann [15], on the other hand, filtered two IMU sensors on the upper-arm and forearm to approximate the orientation of the arm. Similarly, Atrsaei et al. [35] used two IMU sensors on the upper-arm and forearm along with UKF to approximate the pose of the arm. Wang et al. [36] used an IMU to estimate the angular pose of the forearm and EMG to sense the state of the fingers. With such a configuration, along with natural language processing, a user can teach a robot hand-over tasks. However, the methods were focused on approximating the current pose of the arm and did not consider future arm trajectories. In addition, the ability of the above methods to generalize to various participants has not been demonstrated.

Some work has address the classification of future tasks conducted by the user [37]. For instance, three IMU sensors were used along the arm of the user to recognize a task (e.g., tying shoe laces or putting on socks) for controlling a prosthetic arm and using the Long Short Term Memory (LSTM) [38]. A similar approach was taken to classify future activities [39]. The prediction of reaching motion was addressed in [40] while using EMG and gaze data with a Kalman filter. A similar setting was demonstrated by Tortora et al. [41] while combining IMU and EMG sensors to classify between several discrete targets. In a work more related to ours, two IMU's were combined with EMG and Mechanomyography (MMG) to predict future targets among several discrete ones [42]. However, the reaching was conducted only on a table with short trajectories to the targets and while the user is sitting down. Also, the current pose of the human arm is not considered such that the robot can plan a collision-free path. In addition, none of the approaches have demonstrated generability to new users.

In this paper, we aim to rely solely on a low-cost wearable device directly strapped on the human arm to provide an affordable solution. Achieving vision-free pose estimation and prediction would enable further fusion with vision for better estimation in unstructured or partly-occluded environments. Therefore, we explore the sole use of an IMU for learning reaching motions of a human arm. With two IMU sensor located on the upper-arm and forearm, we observe the required data and learning model to predict the target of the reaching arm at early stages of motion in real-time. A learning approach is proposed based on the LSTM model. We investigate the use of raw data from the IMU sensors along with temporal pose predictions of the arm acquired from a secondary ANN model. Therefore, our

proposed method does not require integrating the measured accelerometers in order to predict the position. Integration is often subject to drift and sensitive to typical noisy signals [34]. Contrary to the common approach, our method does not require the use of filters nor early calibration of the wearable device. Hence, it can be used immediately upon placement. The proposed approach estimates the current pose of the human arm and uses it as additional information to predict the reaching target. The current pose of the human arm can also be used by the robot to plan a collision-free path. The approach is shown to be robust to the uncertainties in the positioning of the user and in different body poses. We also observe the robustness of the method to taking-off and re-positioning the device on the arm, and for several users. Unlike prior work, we demonstrate and evaluate the interaction with a robotic arm in human reaching motions.

The main contribution is an approach that enables real-time target prediction in reaching motions of a human arm using affordable, lightweight and easy to use hardware, and without the need for visual perception. With early real-time information, the robot could promptly plan and initiate a response motion in order to interact in a timely manner. The approach is also shown to be able to cope with variance in the pose of the human user standing in front of the robot. While we focus on collaboration with robotic arms, the approach can also be applicable to interactions with prosthetic hands, drones, teleoperation and virtual reality.

The remainder of this paper is organized as follows. Section II establishes the addressed problem, the proposed system and learning method. Furthermore, Section III presents the data collection, experiments, analysis and demonstration for reaching prediction. Finally, Section IV concludes the paper.

II. METHOD
A. PROBLEM FORMULATION
A human user stands in front of a designated interaction-space $\mathcal{W} \in \mathbb{R}^3$. Interaction-space $\mathcal{W}$ is shared with a robotic arm such that both user and robot can reach all points within it (Fig. 1). Let the position of the human wrist at time t be $\mathbf{p}(t) \in \mathbb{R}^3$ relative to some coordinate frame on $\mathcal{W}$. At time $t = 0$, the wrist is at some arbitrary pose. The initial pose is randomly selected close to the body of the user (e.g., side of the hip or palm on the chest). The user then moves his arm to reach some point $\mathbf{p}(t_f) \in \mathcal{W}$ where $t_f < T$ for a pre-defined upper-bound T.

Let $\mathbf{x}(t) \in \mathbb{R}^n$ be the state vector of some n measurable features on the human arm and denote $\tilde{\mathbf{p}}(t) \in \mathbb{R}^3$ to be the estimated position of the wrist at time t. In addition, we denote $\tilde{\mathbf{p}}_f(t) \in \mathcal{W}$ to be the estimated destination of the wrist approximated at time $t < T$. We set two goals. First, we aim to localize the human wrist in real-time during motion. Hence, we search for a map $\Gamma : \mathbb{R}^n \rightarrow \mathbb{R}^3$ that approximates the wrist position $\tilde{\mathbf{p}}(t) = \Gamma(\mathbf{x}(t))$ at time t. Second, we aim to learn a map $\Phi : \mathcal{C} \rightarrow \mathcal{W}$ that provides
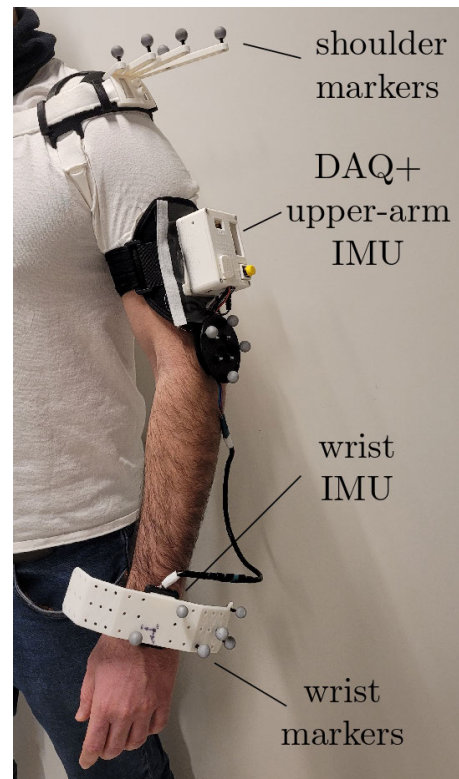
a prediction at time t of the expected wrist target at some time t_f. $\mathcal{C}$ is some unknown space based on measurable features in $\mathbf{x}(t)$ that can provide accurate prediction. In this work, we explore the formulation of $\mathcal{C}$ to acquire a sufficient representation of $\tilde{\mathbf{p}}_f(t) = \Phi(\mathbf{d}(t))$ for $\mathbf{d}(t) \in \mathcal{C}$.

B. SYSTEM
We have designed and fabricated an experimental wearable device consisting of three main components: (a) wrist band with IMU and reflective markers, (b) upper-arm band with IMU and (c) a data acquisition (DAQ) system based on an Arduino Uno board. We have also included shoulder and upper-arm reflective markers for analysis and verification while not used in the modeling. The system is seen in Fig. 2. Each IMU provides 3-axes of acceleration, angular change rate and orientation using accelerometer, gyroscope and a magnetometer, respectively. Hence, two IMU sensors provide a state vector $\mathbf{x}$ with maximum size of $n = 18$.

Measurements of a magnetometer are dependent on the relative orientation to earth's magnetic field. Without loss of generality, we consider an interaction-space at the same orientation relative to the standing position of the human. Hence, we assume that all recordings are taken within a limited orientation range of the human torso relative to the interaction space. Nevertheless, simply having another IMU on the human torso, magnetometer in particular, would

enable relative measurements and additional interaction-spaces around the user.

The acquisition configuration provides real-time data stream of all the given sensors in a frequency of 60 Hz. The reflective markers are tracked using a motion capture system and provide the positions of the shoulder, upper-arm and wrist bands. We note that the marker set is used only for data collection and validation while not required in the eventual system usage. Hence, the described system is composed of low-cost and light-weight (325 g) hardware which is appealing and suited for easy arm movements.

C. DATA COLLECTION AND FORMATION

Data is collected over a set of K episodes. In each episode, the user starts from wrist position $\mathbf{p}(0)$ and moves the arm to final wrist position $\mathbf{p}(t_f) \in \mathcal{W}$. During motion of episode j, sets of states $\mathcal{X}_j = \{\mathbf{x}_j(0), \ldots, \mathbf{x}_j(t_f)\}$ and wrist positions $\mathcal{P}_j = \{\mathbf{p}_j(0), \ldots, \mathbf{p}_j(t_f)\}$ are recorded. For training model Γ, data from all episodes is pre-processed to have a labeled training dataset $\mathcal{M} = \{(\mathbf{x}_i, \mathbf{p}_i)\}_{i=1}^{N}$. The pre-processing and formation of the training data for target prediction will be discussed later. In this work, we wish to explore the ability of data from a single user to provide accurate models. It is assumed that motion patterns of different users are sufficiently similar and further evaluate this in experiments. Hence, data is collected on a single user and later evaluated on other users.

D. WRIST POSITION MODEL

Let θ_{elv} and θ_{yaw} be elevation and yaw angles of the upper-arm, respectively. Similarly, let ϕ_{elv} and ϕ_{yaw} be the forearm elevation and yaw angles, respectively. The position of the wrist $\mathbf{p} = (p_x, p_y, p_z)^T$ is acquired by forward kinematics as given by Soechting and Flanders [43]:

$$p_x = l_u \sin\theta_{elv} \sin\theta_{yaw} + l_f \sin\phi_{elv} \sin\phi_{yaw} \quad (1)$$

$$p_y = l_u \sin\theta_{elv} \cos\theta_{yaw} + l_f \sin\phi_{elv} \cos\phi_{yaw} \quad (2)$$

$$p_z = -l_u \cos\theta_{elv} + l_f \cos\phi_{elv}. \quad (3)$$

where l_u and l_f are the lengths of the upper-arm and forearm, respectively. Assuming l_u and l_f are known, expressions (1)-(3) show that the position of the wrist can be acquired by measuring the orientations of the upper-arm and forearm. While these orientations can be measured by the accelerometers of the two IMU's when in a static pose, the two magnetometers can also do so during arm motion. Nevertheless, acceleration can add viable information for reaching $\mathcal{W}$.

Given training dataset $\mathcal{M}$, we train a feed-forward ANN to obtain map $\tilde{\mathbf{p}}_j = \Gamma_\theta(\mathbf{x}_j)$. Vector θ consists of the trained parameters of the model. In such case, user lengths l_u and l_f are embedded in $\mathcal{M}$ and explicit measurements are not required. Furthermore, the features to be included in state $\mathbf{x}_j$ and its size n such that Γ_θ achieves highest accuracy would be analyzed in the experimental section.

E. LONG-SHORT TERM MEMORY (LSTM)

LSTM is a class of Recurrent Neural-Networks (RNN) aimed to learn sequential data [44]. RNN utilizes previous outputs as inputs while including hidden states. For each time step t, the hidden state vector $\mathbf{h}(t)$ and the output $\mathbf{y}(t)$ are expressed as

$$\mathbf{h}(t) = g_1(W\mathbf{h}(t-1) + U\mathbf{x}(t) + \mathbf{b}_h) \quad (4)$$

and

$$\mathbf{y}(t) = g_2(V\mathbf{h}(t) + \mathbf{b}_y) \quad (5)$$

where W, U, V are weight matrices and $\mathbf{b}_h, \mathbf{b}_y$ are bias vectors. g_1 and g_2 are activation functions. The standard RNN is usually not capable of handling long intervals where back-propagating errors tend to vanish or explode [45]. LSTM, on the other hand, is capable of learning long-term dependencies by utilizing memory about previous inputs for an extended time duration [46]. Along with an hidden state vector, LSTM maintains a cell state vector $\mathbf{c}(t)$. At each time step, the process may choose to read from $\mathbf{c}(t)$, write to it or reset the cell using an explicit gating mechanism. Each cell unit has three gates of the same shape. The input gate controls whether the memory cell is updated or, in other words, which information will be stored. An LSTM cell can be formulated with the following expressions:

$$i(t) = \sigma(W_i[\mathbf{h}(t-1), \mathbf{x}(t)] + \mathbf{b}_i) \quad (6)$$

$$f(t) = \sigma(W_f[\mathbf{h}(t-1), \mathbf{x}(t)] + \mathbf{b}_f) \quad (7)$$

$$o(t) = \sigma(W_o[\mathbf{h}(t-1), \mathbf{x}(t)] + \mathbf{b}_o) \quad (8)$$

w here W_i, W_f and W_o are weight matrices. $\mathbf{b}_i$, $\mathbf{b}_f$ and $\mathbf{b}_o$ are bias vectors. Forget gate $f(t)$ controls whether the memory cell is reset and removes irrelevant information from the cell state. Similarly, output gate $o(t)$ controls whether the information of the current cell state is made visible and adds useful information to the cell state. Both gates have a Sigmoid activation function σ. To modify the cell state, another vector $\tilde{\mathbf{c}}(t)$ is defined as

$$\tilde{\mathbf{c}}(t) = \tanh(W_c[\mathbf{h}(t-1), \mathbf{x}(t)] + \mathbf{b}_c) \quad (9)$$

where W_c and $\mathbf{b}$ are weight matrix and bias vector, respectively. The hyperbolic activation function tanh distributes gradients and, therefore, prevents vanishing or exploding gradients, and allows a cell state information to flow longer. Vector $\tilde{\mathbf{c}}(t)$ is a new candidate that can be applied to the cell state in case the forget state chooses to reset. Hence, the new cell state $\mathbf{c}(t)$ is updated with

$$\mathbf{c}(t) = f(t)\mathbf{c}(t-1) + i(t)\tilde{\mathbf{c}}(t). \quad (10)$$

Furthermore, the hidden state is updated according to

$$\mathbf{h}(t) = \tanh(\mathbf{c}(t)) \cdot o(t). \quad (11)$$

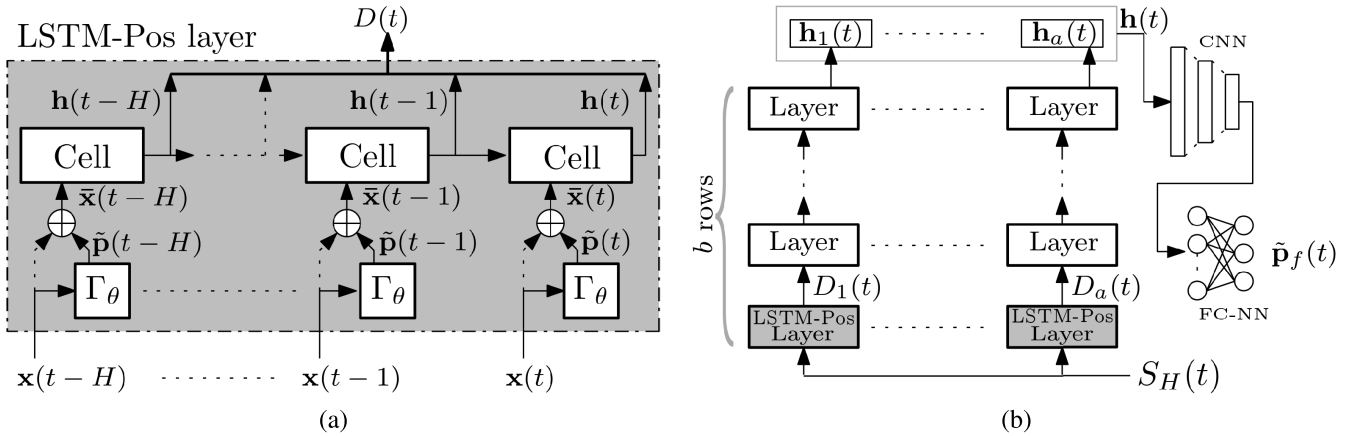The LSTM is trained using recorded data sequences with back-propagation.

FIGURE 3. (a) A scheme of an LSTM-Pos layer receiving a state sequence $S_H(t) = \{\mathbf{x}(t-H), \ldots, \mathbf{x}(t)\}$ of length H with output vector $\mathbf{h}_j(t)$. The requirement for including the raw data with concatenation (denoted with symbol $\oplus$) would be further analysed. (b) The proposed LSTM-Pos network with stacked layers followed by a feed-forward output layer. The network outputs the predicted target of the wrist $\tilde{\mathbf{p}}_f(t)$ at time t.

F. TARGET PREDICTION MODEL

Prediction based on acceleration data requires the integration of the signals along with position information. However, the motion is non-linear and an analytical representation is not available. Consequently and in addition to the sequential nature of the data, we require learning the motion pattern over some period. As noted above, LSTM has the ability to model sequences by selectively remembering certain patterns over some period and learn long-term dependencies. Therefore, we utilize LSTM to explore the various feature and data size requirements in order to predict targets.

Let $S_H(t) \in \mathbb{R}^n \times \ldots \times \mathbb{R}^n$ be the sequence of H past states up to time t given by

$$S_H(t) = \{\mathbf{x}(t-H), \ldots, \mathbf{x}(t-1), \mathbf{x}(t)\}.$$

From each recorded episode $\{\mathcal{X}_j, \mathcal{P}_j\}$, we extract sequences $S_H(H), S_H(H+1) \ldots S_H(t_f)$ and their respected target label $\mathbf{p}(t_f)$. Label $\mathbf{p}(t_f)$ is the last component of $\mathcal{P}_j$. Consequently, we acquire a training dataset from all episodes $\mathcal{L} = \{(S_{H,i}, \mathbf{p}_{f,i})\}_{i=1}^M$ where $S_{H,i}$ and $\mathbf{p}_{f,i} \in \mathcal{W}$ are the ith sequence and corresponding target label, respectively, in the dataset.

Using dataset $\mathcal{L}$, we can directly train an LSTM model for map Φ to predict wrist target. In such case, the input to the LSTM would be a sequence of states where each state has a maximum dimension of $n = 18$. In the experimental section, we will further investigate the importance of IMU features to the accuracy of the prediction. Nevertheless, we hypothesize that the corresponding estimation (using trained model Γ_θ) of wrist positions included in the state sequence would provide viable information for better accuracy. This is analogous to including initial conditions when integrating acceleration. Hence, we consider two alternatives for a new state input $\bar{\mathbf{x}}(t-k)$ to the LSTM. In the first, we would feed only the approximated wrist positions to the LSTM

$$\bar{\mathbf{x}}(t-k) = \Gamma_\theta(\mathbf{x}(t-k))^T. \qquad (12)$$

Alternatively, we observe the concatenation of the wrist position and the original raw data. Hence, for each state in input sequence $S_H(t)$, we concatenate the approximated wrist position to generate a new state $\bar{\mathbf{x}}(t-j)$ given by

$$\bar{\mathbf{x}}(t-k) = \left(\mathbf{x}^T(t-k), \Gamma_\theta(\mathbf{x}(t-k))^T\right)^T. \qquad (13)$$

Consequently, the input to the LSTM will now be the sequence along with the corresponding approximated wrist position. We would analyze these two alternatives in the experimental section.

We denote LSTM with wrist position information as *LSTM-Pos*. The architecture of an LSTM-Pos layer is seen in Fig. 3a. The original state sequence $S_H(t)$ is the input to the layer followed by concatenation (13) using trained model Γ_θ. When considering state representation (12), the concatenation does not occur and approximated wrist positions are fed directly into the cells. A cell in the layer receives vector $\bar{\mathbf{x}}(t-k)$ and outputs hidden state vector $\mathbf{h}(t-k) \in \mathbb{R}^m$ where m is an hyper-parameter. Hidden state vectors are passed between the cells and additionally collected to hidden state sequence $D(t) = \{\mathbf{h}(t-H), \ldots, \mathbf{h}(t)\}$. Sequence $D(t)$ is the output of the LSTM-Pos layer.

Fig. 3b illustrates the entire LSTM-Pos network. LSTM-Pos has b rows while each row has a layers. a and b are hyper-parameters of the network. Only the layers of the first row are LSTM-Pos layers where each outputs an hidden state sequence $D_j(t)$ for $j = 1, \ldots, a$. The other layers are regular LSTM layers that receive hidden state sequences and output updated ones. While all layers in a row have similar architecture, they include a random dropout such that $D_i(t) \neq D_j(t)$ for any $i, j \in \{1, \ldots, a\}$ and $i \neq j$. Furthermore, only hidden states at time t are outputted from the last row. Therefore, a hidden states $\mathbf{h}_1(t), \ldots, \mathbf{h}_a(t)$ are used and are the input to a standard Convolutional Neural-network (CNN) and further to a fully-connected NN (FC-NN). The FC-NN outputs the approximated target wrist position $\tilde{\mathbf{p}}_f(t)$ predicted at time t. In preliminary testings, training the model

without a CNN failed to converge and is, therefore, essential. In the experimental section, we compare the LSTM-Pos to the standard LSTM which has the same architecture while not including wrist position information.

G. CURRICULUM LEARNING

Preliminary testing has shown that training the models directly with all data may lead to predictions converging to the geometric center of the workspace. Hence, we use *Curriculum Learning* (CL) [47]. The common approach is to introduce the model with data in an organized order from easy samples to hard ones. Such approach commonly provides better performance than random data shuffling.

We use preliminary insights on model performance to tackle the problem of convergence to a local minima. We propose to take a reverse CL approach and train with harder samples first. In early studies, we have noticed that it is harder to predict the wrist pose and target at early stages of the reaching motion. This is due to low data variance when initiating motion. Hence, we train a model with N_{cl} batches along the motion starting from time $t = 0$. Once the loss value of the model successfully reaches below some user-defined value γ_{cl}, the next batch of further time frame is added to the training. When optimizing the hyper-parameters of some model, if value γ_{cl} is not reached for some batch, the model is disqualified.

III. EXPERIMENTS

We use the experimental system described in Section II-B to collect data and analyze the proposed approach. Collection was performed by one human participant (participant 1) with the system on his left arm. Arm lengths can be seen in Table 3. To achieve uniform and guided data collection, a rectangular collection board was mounted in front of the participant as seen in Fig. 4. The board was equally divided into 42 squares. The participant was asked to start from arbitrary arm locations near the body and reach each of the squares several times. In each square, sized 110 mm, the user could touch any random location as desired. To acquire a robust model, the participant also removed the system and re-positioned it on the arm several times during recordings. It is important to note that the proposed method processes raw data and uses relative IMU signals. Therefore, calibration of the IMU sensors is not required and the device can be used immediately upon positioning.

A total of $K = 840$ reaching episodes were collected with 20 episodes for each square. Preliminary analysis has shown that natural reaching motions in the described setup are no longer than 2 seconds. Hence, each episode was recorded in 60 Hz for $T = 2$ *sec* leading to 120 samples for an episode. Therefore, a set of $N = 100,800$ samples are available for training a wrist position model. The participant recorded reaching tasks with various torso orientations including facing the collection board, body perpendicular to the board and while sitting down. Hence, a large distribution of shoulder motions during arm reaching can be seen in

TABLE 1. Approximation accuracy of wrist position during reaching motion.

	w/o CL (mm)	w/ CL (mm)
All sensors	58.24 ± 4.79	48.62 ± 5.07
Accelerometers	99.71 ± 18.43	63.78 ± 14.73
Accel. & Magn.	76.1 ± 5.63	58.81 ± 5.83
Gyroscopes	144.08 ± 17.24	124.79 ± 20.32
Magnetometers	93.79 ± 11.87	65.55 ± 8.43
Wrist only	64.51 ± 5.19	55.76 ± 5.86
Upper-arm only	149.1 ± 14.39	59.78 ± 10.8

Fig. 5 showing high variance perturbations. Note however that explicit information about the pose of the human during reaching is not included in the training data.

A large testing dataset was also collected independent of the training set and includes 336 episodes when reaching eight episodes for each square in various torso poses. The test set also included removing and re-positioning the device between episodes in order to test robustness. We next evaluate the training wrist position and target prediction models with the data. The datasets generated during and analysed during the current study are available in a designated repository[1]

A. WRIST POSITION PREDICTION EVALUATION

We now analyse the wrist position model Γ as discussed in Section II-D. We analyze the accuracy of various ANN models trained with different features in the data. We test the ability of a model to predict the wrist position if only some of the features are available including: only accelerometers, only magnetometers, only gyroscopes, only accelerometers and magnetometers, only wrist IMU band, only upper-arm IMU band or all available sensors. For each variation, we optimized the hyper-parameters of the ANN to minimize the RMSE loss function along with the Adam optimizer. The Optuna framework [48] was used for optimization of the number of layers and hidden neurons. For example, the optimal architecture for the model with all sensing features included three hidden layers with 512 neurons each and a rectified linear unit (ReLU) activation function. We also compare between direct and CL training of the ANN. For the CL, we manually optimized the batch size and loss threshold yielding $N_{cl} = 10$ and $\gamma_{cl} = 58$ *mm*, respectively.

Table 1 summarizes the accuracy results for all variations. It is clear that CL provides accuracy improvement in all feature representations. Magnetometers (denoted *Magn.* in the table) by themselves supposedly should provide enough information to localize the arm. However and due to variations in torso orientations, moderate accuracy is provided with only magnetometers. Only using accelerometers (denoted *Accel.* in the table) reaches similar accuracy as they are able to encapsulate motion flow patterns but lack orientation information. Evidently, combining accelerometers and magnetometers provides accuracy improvement. Furthermore, using all available sensors of both IMU sensors

---

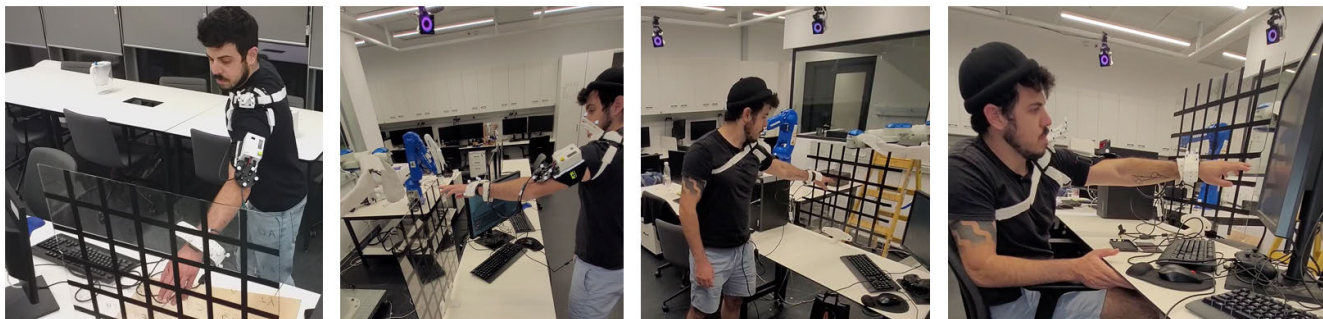[1]Dataset available at: https://bit.ly/3COveC8

FIGURE 4. Participant 1 collecting data of arm reaching in various torso orientations including facing the collection board, body perpendicular to the board and while sitting down.

TABLE 2. Target prediction accuracy for LSTM and LSTM-Pos.

Features	LSTM		LSTM-Pos	
	Mean (mm)	Std. (mm)	Mean (mm)	Std. (mm)
All sensors	194.63	73.07	61.46	11.26
Accelerometers	195.42	70.03	66.27	12.42
Accel. and Magn.	195.37	69.96	65.53	12.56
Gyroscopes	195.41	70.02	71.58	14.04
Magnetometers	195.43	69.99	72.71	19.12
Approx. wrist position	-	-	59.99	12.01
Wrist only	195.39	69.99	67.56	13.9
Upper-arm only	195.42	70.03	69.04	17.26

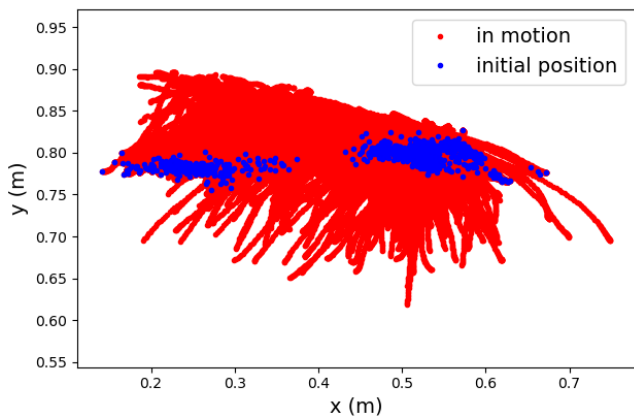**FIGURE 5.** Shoulder positions relative to the board (at $y = 0$) before initiating motion (blue) and during motions (red).

FIGURE 6. Target prediction error (mean and standard deviation) with respect to the number of the past states H.

(including the gyroscopes) provides the lowest error. The results also show that having only one IMU either on the upper-arm or the wrist could provide fairly accurate wrist position. We note that these results are similar to results reported using the Unscented Kalman filter with an RMS error of 49 mm [35].

B. TARGET PREDICTION EVALUATION

With the above NN models for approximating wrist position, we analyze the prediction of the target in reaching motions with the LSTM-Pos. We compare LSTM-Pos to an LSTM without including explicit approximation of wrist positions. Here also, we analyze the required features necessary to achieve accurate prediction. The hyper-parameters of models
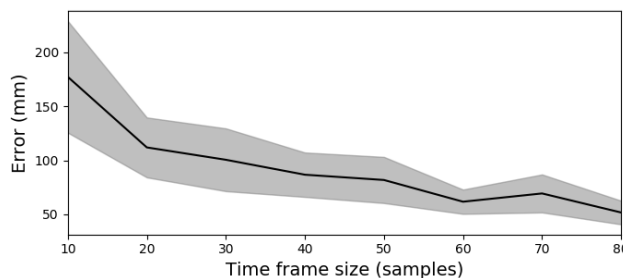
with different feature variations were optimized using the Optuna framework to yield the lowest RMSE loss value. For example, the optimal hyper-parameters for the model with all sensing features include LSTM with $a = 256$ and $b = 2$, $m = 64$, CNN with three convolutional layers and FC-NN with one hidden layer of 14 nodes, yielding 42,291 trainable parameters. In this part, all training is performed with CL since it provides better learning as demonstrated previously.

We initially observed the required number of past states H to acquire an accurate model. Fig. 6 shows the mean prediction error over the test data when including all sensors and with regards to H. Including more information provides, as one would expect, better accuracy. Other feature variations exhibit similar behaviour. Having a large H means that the first prediction would arrive later in the reaching motion. Hence, we face a trade-off between earlier prediction and accuracy. In further analysis we choose to use $H = 60$ as it provides low error along with first prediction at half-way
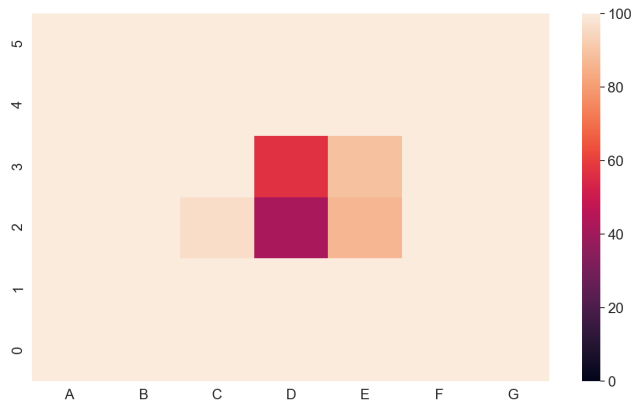
FIGURE 7. Typical heat-map illustrating the mean error (in *mm*) across the collection board when not including wrist position approximation of LSTM-Pos. All models converged their prediction to the center of the board.

FIGURE 8. Target prediction error (mean and standard deviation) of LSTM-Pos with respect to reaching motion time evaluated on all test episodes.

TABLE 3. Arm lengths of all participants.

Participant	1	2	3
l_u (cm)	29	33.5	25
l_f (cm)	28.5	30.3	24.7

through-out the reaching motion. The chosen prediction horizon provides at least one second before the hand reaches the target which is valuable for planning and initiating robot motion. While not implemented in this work, these results show that one could train several models with ascending H. Hence, coarse accuracy would be provided at the beginning of the episode (e.g., $H = 10$) for initial motion of the robot. Then, accuracy would be improved with better models as more information arrive.

Table 2 summarizes the target prediction errors (mean and standard deviation) along the reaching motion of all test data and for different feature variations. First, directly using the data to train an LSTM (without including wrist position approximation) fails to produce feasible predictions. All LSTM models converged to the mean of the target labels. Fig. 7 shows an heatmap illustrating the error distribution across the collection board. Predictions for all test episodes output the center position of the collection board leading to low accuracy performance.

LSTM-Pos, on the other hand, provides much lower prediction errors for all feature variations. In particular, having only the sequence of approximated wrist positions from the NN fed into the LSTM is sufficient to acquire the best model. Including also the raw measurements from all sensors fairly provides the same accuracy. The results also indicate that having only one band can be sufficient for moderate accuracy. Fig. 8 shows the mean error along the test episodes for LSTM-Pos with only approximated wrist positions and when including also raw sensor measurements (all sensors). Similarly, Fig. 9 illustrates the mean errors across the collection board through the motion time. The above results achieve accuracy of approximately less than 70% of an adult human palm breadth (87.5 mm) [49]. In a hand-over task, the user can slightly refine the position of the hand at the last stage of the motion if required. Hence, such an accuracy is sufficient for feasible HRC tasks as would be demonstrated in the next section.
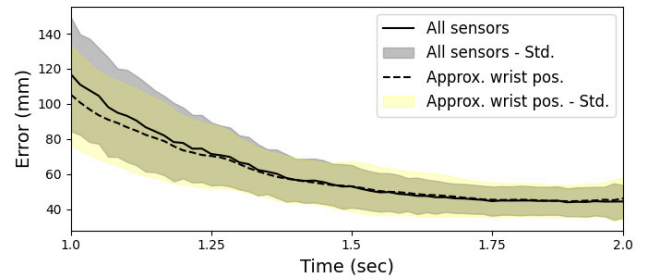
C. HRC DEMONSTRATION

We have conducted an HRC demonstration where a participant reaches his arm towards a robotic arm. During the reaching motion, the robotic arm receives continuous stream of target predictions acquired from the trained model. When acquiring the first prediction, the robot plans motion for rendezvous and initiates motion. As a new prediction arrives, the motion plan of the robot is updated accordingly. Participant 1 along with two additional ones not included in training participated in the demonstration. The participants have different arm lengths as seen in Table 3. They were instructed to stand in front of the robot at desired body poses such that their wrist can reach the shared workspace. These poses are roughly similar to the ones used for collecting training data and within the same distribution. Furthermore, the robot was not given any information regarding the pose of the human participant. The participants were asked to perform 15 reaching motions to arbitrary locations in front of them. They were also instructed to initiate motion from any desired arm pose. We define a successful trial as a one where the robot reached within the vicinity of the human hand with minimal distance of 60 mm. The distance was measured using the motion capture system.

Table 4 provides the rendezvous success rate for the three participants. The success for Participant 1 is the highest since the model is trained on data collected from him. Nevertheless, the two other participants achieved rather high success rates considering that they have different arm lengths and did not contribute data to model training. Hence, the model can be transferred to a new user with relatively good accuracy. The data collection process, which included re-applying the system various times on the arm and the recording of different body poses (evaluated in Fig. 5), enables the LSTM-Pos to overcome uncertainties and learn general motion patterns that are common to different individuals. In addition, the unique architecture and CL training of the LSTM-Pos assist the
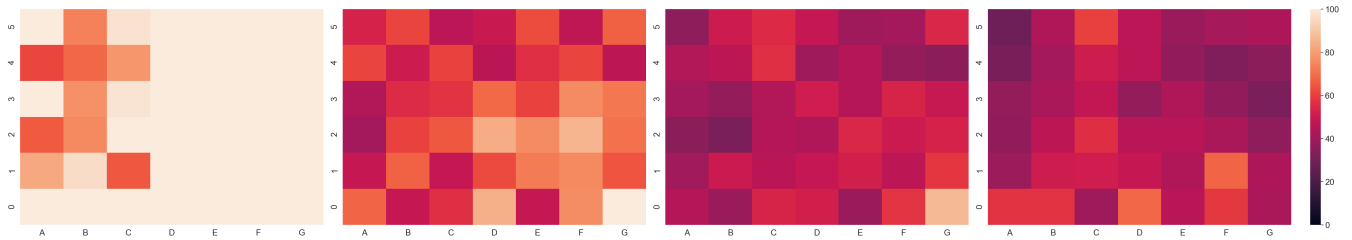
FIGURE 9. Heat-maps illustrating the mean error (in *mm*) across the collection board over time and when considering all sensors. Mean error from left to right: *Error*($t = 1$ *sec*) = 116.98 *mm*, *Error*($t = 1.33$ *sec*) = 62.41 *mm*, *Error*($t = 1.65$ *sec*) = 46.86 *mm* and *Error*($t = 2$ *sec*) = 44.47 *mm*.
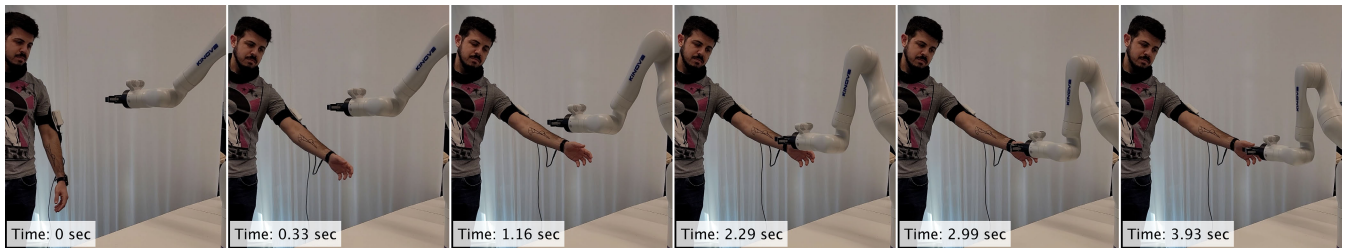

FIGURE 10. Reaching test of participant 1.
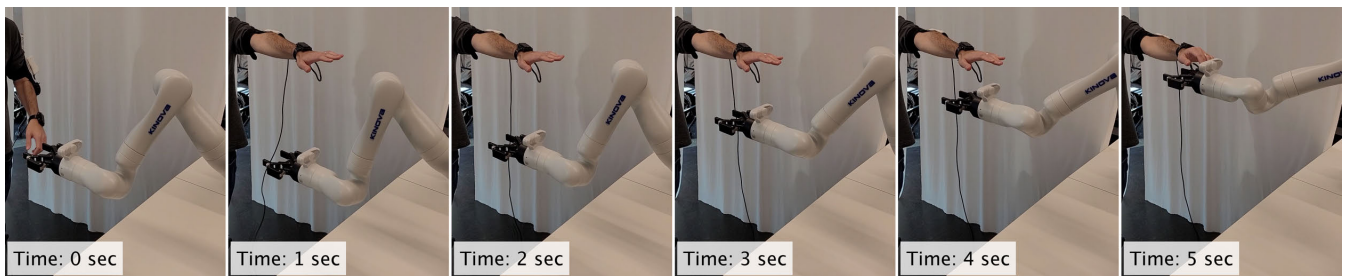
FIGURE 11. Reaching test of participant 3.

TABLE 4. Success rate for reaching experiment.

Participant	Success rate
1	93.3%
2	80.0%
3	86.6%

FIGURE 12. Motion of participant 2 between two arm poses within the interaction space while the robotic arm follows.

learning of vital motion features. Hence, LSTM-Pos was able to provide high performance to the new participants. Also, while the participants were not instructed to stand in a specific pose, yet the model was able to provide accurate predictions. This can be explained by the ability of the model to infer from motion patterns, not only regarding the trajectory of the arm, but also about the human pose. In other words, information regarding human pose is embedded within state trajectories and the model can cope with pose variance for successful HRC. Hence, the model is shown to provide good accuracy within some implicit pose distributions of the human participant.

We note that the robot arm reached the vicinity of the human hand in all trials but with a distance larger than the defined bound. Target wrist prediction was achieved in mid-motion and the robot initiated motion before the participant reached the end of the episode. For safety reasons, the velocity of the robot was maintained low. Hence, the participant was required to shortly wait for the arm. Nevertheless, having the robot move faster would enable faster HRC tasks. Figures 10-12 show demonstration snapshots of some reaching trials for the three participants towards the robotic arm. Participant 3 in Fig. 12 demonstrated the ability of the model to predict motion from various initial poses of the arm. Fig. 13 shows a demonstration of handing over a bottle to the robot using the system. These results show

FIGURE 13. An HRC demonstration in which participant 1 hands-over a bottle to the robot.

that a relatively small dataset of IMU data can be used to train a sufficiently accurate model for feasible predictions of reaching targets in HRC scenarios. All experiments and demonstration videos can be seen in the supplementary video.

IV. CONCLUSION

We have presented the problem of learning human reaching motions and predicting future target locations. We investigated the use of a wearable device composed of two IMU sensors located on the upper-arm and wrist. First, a simple ANN was trained to predict the current position of the wrist. Then, an LSTM-based architecture was proposed where approximation of the wrist position along some time frame was included. Using data collected from a single user, we have tested various features to be included in the model. Results show that having all features from the sensors along with curriculum learning achieve good accuracy for approximating current wrist position. With such model and without additional raw data, we reach sufficient accuracy for early prediction of the target wrist position. Furthermore, we have shown several reaching demonstrations where a robot planned and moved towards the human arm in real-time based solely on information from the wearable device. High success rates were demonstrated from the participant whom collected data and two other ones not included in the training. Hence, a relatively small amount of collected data from one user achieved sufficient accuracy for feasible predictions of reaching targets in HRC scenarios. Once trained, the model and system can be deployed in various spaces with no further effort.

The results have shown some decline in performance when using a single-user trained model for predicting motion of new users. In order to provide a multi-user model, future work may consider fine-tuning the existing model to a new user based on a limited amount of collected data. Alternatively, training data from several users may be used to acquire a robust model. The system could also be fused with additional sensors. For instance, an additional IMU can be added on the human torso to provide information regarding body orientation and upper body shift. Vision can also be integrated in order to provide a complete solutions when a line-of-sight is not continuous. Similarly, one could integrate Force-Myography [50], [51] or EMG [31] to have additional information about the pose of the palm and fingers. An additional system on the other arm could enable the classification of tasks performed by the human towards assistance by a robot.

REFERENCES

[1] G. Rizzolatti, L. Fogassi, and V. Gallese, "Neurophysiological mechanisms underlying the understanding and imitation of action," *Nature Rev. Neurosci.*, vol. 2, no. 9, pp. 661–670, Sep. 2001.

[2] A. Ajoudani, A. M. Zanchettin, S. Ivaldi, A. Albu-Schäffer, K. Kosuge, and O. Khatib, "Progress and prospects of the human–robot collaboration," *Auto. Robots*, vol. 42, no. 5, pp. 957–975, Jun. 2018.

[3] W. Yang, C. Paxton, A. Mousavian, Y. Chao, M. Cakmak, and D. Fox, "Reactive human-to-robot handovers of arbitrary objects," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2021, pp. 3118–3124.

[4] V. Ortenzi, A. Cosgun, T. Pardi, W. P. Chan, E. Croft, and D. Kulic, "Object handovers: A review for robotics," *IEEE Trans. Robot.*, vol. 37, no. 6, pp. 1855–1873, Dec. 2021.

[5] R. Luo, R. Hayne, and D. Berenson, "Unsupervised early prediction of human reaching for human–robot collaboration in shared workspaces," *Auto. Robots*, vol. 42, no. 3, pp. 631–648, Mar. 2018.

[6] P. A. Lasota, T. Song, and J. A. Shah, *A Survey of Methods for Safe Human-Robot Interaction*. Foundations and Trends, 2017.

[7] N. M. Gamage, D. Ishtaweera, M. Weigel, and A. Withana, *So Predictable! Continuous 3D Hand Trajectory Prediction in Virtual Reality*. New York, NY, USA: Association for Computing Machinery, 2021, pp. 332–343.

[8] T. Flash and N. Hogan, "The coordination of arm movements: An experimentally confirmed mathematical model," *J. Neurosci.*, vol. 5, no. 7, pp. 1688–1703, Jan. 1985.

[9] Y. Uno, M. Kawato, and R. Suzuki, "Formation and control of optimal trajectory in human multijoint arm movement," *Biol. Cybern.*, vol. 61, no. 2, pp. 89–101, Jun. 1989.

[10] C. M. Harris and D. M. Wolpert, "Signal-dependent noise determines motor planning," *Nature*, vol. 394, no. 6695, pp. 780–784, Aug. 1998.

[11] O. S. Oguz, Z. Zhou, and D. Wollherr, "A hybrid framework for understanding and predicting human reaching motions," *Frontiers Robot. AI*, vol. 5, p. 27, Mar. 2018.

[12] Y. Tamura, M. Sugi, J. Ota, and T. Arai, "Prediction of target object based on human hand movement for handing-over between human and self-moving trays," in *Proc. ROMAN 15th IEEE Int. Symp. Robot Human Interact. Commun.*, Sep. 2006, pp. 189–194.

[13] Z. Peng, T. Genewein, and D. A. Braun, "Assessing randomness and complexity in human motion trajectories through analysis of symbolic sequences," *Frontiers Human Neurosci.*, vol. 8, p. 168, Mar. 2014.

[14] M. Kohler, "Using the Kalman filter to track human interactive motion–modelling and initialization of the Kalman filter for translational motion," Tech. Rep., 1997.

[15] X. Yun and E. R. Bachmann, "Design, implementation, and experimental results of a quaternion-based Kalman filter for human body motion tracking," *IEEE Trans. Robot.*, vol. 22, no. 6, pp. 1216–1227, Dec. 2006.

[16] D. Wu and L. Shao, "Leveraging hierarchical parametric networks for skeletal joints based action segmentation and recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 724–731.

[17] C. Pérez-D'Arpino and J. A. Shah, "Fast target prediction of human reaching motion for cooperative human–robot manipulation tasks using time series classification," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2015, pp. 6175–6182.

[18] A. M. Zanchettin and P. Rocco, "Probabilistic inference of human arm reaching target for effective human–robot collaboration," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2017, pp. 6595–6600.

[19] Y. Cheng, W. Zhao, C. Liu, and M. Tomizuka, "Human motion prediction using semi-adaptable neural networks," in *Proc. Amer. Control Conf. (ACC)*, Jul. 2019, pp. 4884–4890.

[20] J. Zhao, S. Gong, B. Xie, Y. Duan, and Z. Zhang, "Human arm motion prediction in human–robot interaction based on a modified minimum jerk model," *Adv. Robot.*, vol. 35, nos. 3–4, pp. 205–218, Feb. 2021.

[21] C. T. Landi, Y. Cheng, F. Ferraguti, M. Bonfè, C. Secchi, and M. Tomizuka, "Prediction of human arm target for robot reaching movements," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Nov. 2019, pp. 5950–5957.

[22] H. Ding, G. Reißig, K. Wijaya, D. Bortot, K. Bengler, and O. Stursberg, "Human arm motion modeling and long-term prediction for safe and efficient human-robot-interaction," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2011, pp. 5875–5880.

[23] Z. Liu, Q. Liu, W. Xu, Z. Liu, Z. Zhou, and J. Chen, "Deep learning-based human motion prediction considering context awareness for human–robot collaboration in manufacturing," *Proc. CIRP*, vol. 83, pp. 272–278, 2019.

[24] J. Zhang, H. Liu, Q. Chang, L. Wang, and R. X. Gao, "Recurrent neural network for motion trajectory prediction in human–robot collaborative assembly," *CIRP Ann.*, vol. 69, no. 1, pp. 9–12, 2020.

[25] R. Liu and C. Liu, "Human motion prediction using adaptable recurrent neural networks and inverse kinematics," *IEEE Control Syst. Lett.*, vol. 5, no. 5, pp. 1651–1656, Nov. 2021.

[26] K. Fragkiadaki, S. Levine, P. Felsen, and J. Malik, "Recurrent network models for human dynamics," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 4346–4354.

[27] J. Martinez, M. J. Black, and J. Romero, "On human motion prediction using recurrent neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4674–4683.

[28] T. Salzmann, B. Ivanovic, P. Chakravarty, and M. Pavone, "Trajectron++: Dynamically-feasible trajectory forecasting with heterogeneous data," in *Computer Vision—ECCV 2020*, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds. Cham, Switzerland: European Computer Vision Association, 2020, pp. 683–700.

[29] Q. Wu, G. Xu, M. Li, L. Chen, X. Zhang, and J. Xie, "Human pose estimation method based on single depth image," *IET Comput. Vis.*, vol. 12, no. 6, pp. 919–924, Sep. 2018.

[30] A. Toshev and C. Szegedy, "DeepPose: Human pose estimation via deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1653–1660.

[31] L. Bi, A. G. Feleke, and C. Guan, "A review on EMG-based motor intention prediction of continuous human upper limb motion for human–robot collaboration," *Biomed. Signal Process. Control*, vol. 51, pp. 113–127, Jul. 2019.

[32] Y. Nakanishi, T. Yanagisawa, D. Shin, R. Fukuma, C. Chen, H. Kambara, N. Yoshimura, M. Hirata, T. Yoshimine, and Y. Koike, "Prediction of three-dimensional arm trajectories based on ECoG signals recorded from human sensorimotor cortex," *PLoS ONE*, vol. 8, no. 8, Aug. 2013, Art. no. e72085.

[33] J. A. Corrales, F. A. Candelas, and F. Torres, "Hybrid tracking of human operators using IMU/UWB data fusion by a Kalman filter," in *Proc. 3rd ACM/IEEE Int. Conf. Human-Robot Interact. (HRI)*, Mar. 2008, pp. 193–200.

[34] Y. Tian, X. Meng, D. Tao, D. Liu, and C. Feng, "Upper limb motion tracking with the integration of IMU and Kinect," *Neurocomputing*, vol. 159, pp. 207–218, Jul. 2015.

[35] A. Atrsaei, H. Salarieh, A. Alasty, and M. Abediny, "Human arm motion tracking by inertial/magnetic sensors using unscented Kalman filter and relative motion constraint," *J. Intell. Robotic Syst.*, vol. 90, nos. 1–2, pp. 161–170, May 2018.

[36] W. Wang, R. Li, Y. Chen, Y. Sun, and Y. Jia, "Predicting human intentions in human–robot hand-over tasks through multimodal learning," *IEEE Trans. Autom. Sci. Eng.*, vol. 19, no. 3, pp. 2339–2353, Jul. 2022.

[37] J.-W. Cui, Z.-G. Li, H. Du, B.-Y. Yan, and P.-D. Lu, "Recognition of upper limb action intention based on IMU," *Sensors*, vol. 22, no. 5, p. 1954, Mar. 2022.

[38] J. Cui and Z. Li, "Prediction of upper limb action intention based on long short-term memory neural network," *Electronics*, vol. 11, no. 9, p. 1320, Apr. 2022. [Online]. Available: https://www.mdpi.com/2079-9292/11/9/1320

[39] I. E. Jaramillo, C. Chola, J.-G. Jeong, J.-H. Oh, H. Jung, J.-H. Lee, W. H. Lee, and T.-S. Kim, "Human activity prediction based on forecasted IMU activity signals by sequence-to-sequence deep neural networks," *Sensors*, vol. 23, no. 14, p. 6491, Jul. 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/14/6491

[40] N. E. Krausz, D. Lamotte, I. Batzianoulis, L. J. Hargrove, S. Micera, and A. Billard, "Intent prediction based on biomechanical coordination of EMG and vision-filtered gaze for end-point control of an arm prosthesis," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 6, pp. 1471–1480, Jun. 2020.

[41] S. Tortora, S. Michieletto, F. Stival, and E. Menegatti, "Fast human motion prediction for human–robot collaboration with wearable interface," in *Proc. IEEE Int. Conf. Cybern. Intell. Syst. (CIS) IEEE Conf. Robot., Autom. Mechatronics (RAM)*, Nov. 2019, pp. 457–462.

[42] S. Yang, N. P. Garg, R. Gao, M. Yuan, B. Noronha, W. T. Ang, and D. Accoto, "Learning-based motion-intention prediction for end-point control of upper-limb-assistive robots," *Sensors*, vol. 23, no. 6, p. 2998, Mar. 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/6/2998

[43] J. F. Soechting and M. Flanders, "Errors in pointing are due to approximations in sensorimotor transformations," *J. Neurophysiology*, vol. 62, no. 2, pp. 595–608, Aug. 1989.

[44] Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: LSTM cells and network architectures," *Neural Comput.*, vol. 31, no. 7, pp. 1235–1270, Jul. 2019.

[45] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, Mar. 1994.

[46] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[47] X. Wang, Y. Chen, and W. Zhu, "A survey on curriculum learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 9, pp. 4555–4576, Sep. 2022.

[48] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2019.

[49] N. K. Bayraktar and E. T. Özşahin, "Anthropometric measurement of the hand," *Eastern J. Med.*, vol. 23, no. 4, pp. 298–301, 2018.

[50] N. D. Kahanowich and A. Sintov, "Robust classification of grasped objects in intuitive human–robot collaboration using a wearable force-myography device," *IEEE Robot. Autom. Lett.*, vol. 6, no. 2, pp. 1192–1199, Apr. 2021.

[51] E. Bamani, N. D. Kahanowich, I. Ben-David, and A. Sintov, "Robust multi-user in-hand object recognition in human–robot collaboration using a wearable force-myography device," *IEEE Robot. Autom. Lett.*, vol. 7, no. 1, pp. 104–111, Jan. 2022.

NADAV D. KAHANOWICH received the B.S. and M.S. degrees in mechanical engineering from Tel Aviv University, in 2020 and 2023, respectively. His current research interests include artificial intelligence, robotics, and mechanical design.

AVISHAI SINTOV (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees from the Department of Mechanical Engineering, Ben-Gurion University of the Negev, in 2008, 2012, and 2016, respectively. He is currently an Assistant Professor with the School of Mechanical Engineering, Tel Aviv University. Previously, he was a Postdoctoral Fellow with the University of Illinois at Urbana–Champaign, from 2016 to 2018, and Rutgers University, from 2018 to 2019. His research interests include human–robot collaboration, dynamic manipulations, grasping, regrasping, constrained systems, machine learning, and robot design.

● ● ●