Dynamic Model Informed Human Motion Prediction Based on Unscented Kalman Filter

Wansong Liu[1], Xiao Liang[2], Minghui Zheng[1]

***Abstract*—Human motion prediction is the foundation stone of human-robot collaboration (HRC) in intelligent manufacturing. The non-linear and stochastic nature of human motion has made it challenging to predict the motion accurately. Many recent deep-learning-based approaches, e.g., convolutional neural networks or recurrent neural networks (RNNs) have been applied to address this challenge. On the other hand, existing works tend to ignore the importance of human dynamics in motion prediction, especially, the effect of muscle force on the motion. This paper proposes a novel dynamic model informed motion prediction method. It utilizes an unscented Kalman filter (UKF) to predict the state of the future arm dynamic model such that the future motion of the human arm can be obtained. In particular, the arm dynamic model is developed based on Lagrangian mechanics and represented by differential equations. Embracing the future muscle force predicted by RNN into the differential equations, such a dynamic model is capable of explicitly establishing the intrinsic relation between the future muscle force and the corresponding future arm motion. UKF is leveraged to predict the future joint position and velocity of the human arm based on the dynamic model. Experiments on three motion datasets validate that the proposed prediction method, compared to the traditional RNN-based prediction using skeleton vectors, significantly improves the prediction accuracy regarding elbow and wrist positions.**

***Index Terms*—Human motion prediction, Human-robot collaboration, Unscented Kalman filter**

I. Introduction

In the past decade, great attention has been paid to HRC [1]–[3]. In a seamless and safe collaboration, (e.g., a robot arm delivers tools to a human worker when doing an assembly task), the collaborators need to recognize each other's motion and take possible future motion into account such that the efficiency and safety can be guaranteed in a confined workspace. Therefore, to promote efficient and safe HRC, extensive efforts have been made on human motion prediction.

Early tracking-based approaches rely on probabilistic models to track and predict human motion using particle filter (PF) [4], [5]. Another type of methods formulates the human motion prediction as an optimization problem with respect to certain metrics [6], [7]. In addition, human motions are encoded into some latent-variable models, e.g., Gaussian processes [8], restricted Boltzmann machine [9], and hidden Markov models [10], in another category of approaches. More recently, neural networks have shown the great capability of handling complicated motion prediction [11]–[13]. Among the various neural networks, RNN with different architectures, e.g., long short-term memory (LSTM) [14] and gated recurrent units (GRU) [15], is widely applied to capture the dependencies in the motion data since the RNN hidden states can store the past motion characteristics [16]. Moreover, different RNN-based works were developed to improve the prediction performance. For example, Liu et al. [17] first used RNN with LSTM to predict the wrist motion, and then leveraged inverse kinematics to obtain the whole arm's future motion. Using LSTM as the recurrent layer, an encoder-recurrent-decoder network was applied to learn and forecast human kinematics in videos [18]. A sequence-to-sequence residual architecture utilizing GRU showed the effectiveness in both short-term and long-term motion predictions [19].

Despite RNN having the advantage of learning motion contexts, it still has several drawbacks. First, RNN is difficult to train since the long-term components grow exponentially or convert to norm 0 easily [20]. Moreover, existing works, e.g., [17], [18], applied N-to-1 structure to predict the motion of the next step iteratively, which tends to accumulate errors throughout the entire predicted sequence. The last critical problem is that most RNN-based works fail to build the intrinsic connection between motion and the human physical body.

To incorporate the physical information of the human body into network training, several approaches utilize Lie algebra to represent separate kinematic chains of the human body, such that anatomical constraints during the motion are encoded naturally [21]. For example, a special Euclidean group SE(3) was used to indicate the relative translation and rotation of each skeleton in [21]. In addition, to handle the inconsistent bone length issue caused by bringing the relative translation into training (e.g., [22] and [23]), with the bone length fixed, the relative rotation of each skeleton was represented using a special orthogonal group SO(3). In brief, instead of using the traditional Euclidean training loss, the mentioned Lie-algebra-based approaches inherently use the geodesic training loss such that the relative geometric structure of the human skeleton is enhanced during the network training.

Even though the approaches using SO(3) or SE(3) obtain impressive performances in human motion prediction, the Lie-algebra-based approaches still ignore the dynamics of human motion, especially the muscle force which serves as the most

This work was supported by the USA National Science Foundation under Grants No.2026533 and No.1928595.

[1] Wansong Liu and Minghui Zheng are with the Mechanical and Aerospace Engineering Department, University at Buffalo, Buffalo, NY14260, USA. Emails: {wansongl, mhzheng}@buffalo.edu.

[2] Xiao Liang is with the Civil, Structural and Environmental Engineering Department, University at Buffalo, Buffalo, NY14260, USA. Email: liangx@buffalo.edu.

* Correspondence to Minghui Zheng and Xiao Liang.

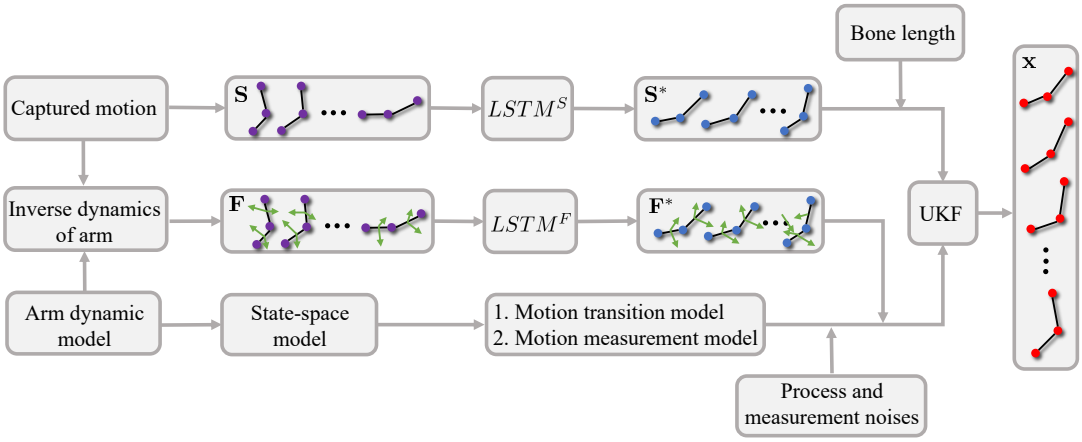


Fig. 1: Overview of proposed prediction approach: green arrows indicate the muscle forces, purple dots represents the joints of the observed poses, blue dots stands for the joints of the predicted poses, and red dots are the predicted joints after UKF.

significant motivation of the human motion. Considering that electromyography (EMG) is capable of directly measuring muscle response [24], the relationship between the motion and muscle force has been extensively explored in EMG-based approaches, e.g., [25]–[27]. However, as discussed in [28], [29], EMG signals contain lots of fuzziness and are easily affected by environmental conditions.

Therefore, using the Lagrangian approach to develop the inverse dynamics of the human body is leveraged as an alternative way to map the human motion to the corresponding muscle forces explicitly. For example, Li et al. [30] used Lagrangian dynamic equations to describe the full-body movement of the person in video frames and to estimate the force of person-object interactions actuated by the human. Lv et al. [31] proposed a data-driven inverse dynamics approach to estimate human joint torques, and the effectiveness of the proposed approach was validated in various human movements. Cao and Nevatia [32] applied inverse dynamics and direct dynamics, respectively, to calculate forces and re-construct occluded or non-captured human poses.

On the other hand, integrating the Lagrangian-mechanics-based dynamic model into human motion prediction remains an ongoing challenge considering that future motion and force are both unknown at the current time instant. To overcome this challenge, this paper proposes a dynamic model informed motion prediction approach, which uses UKF to predict the human motion based on future human dynamics. Specifically, we focus on human arm motion, and apply RNN and Lagrangian mechanics to obtain the future arm dynamic model. Considering that such a dynamic model is highly nonlinear, UFK is leveraged to predict the model state, which is the joint position and velocity. In addition, since it is impossible to use sensors to measure the future arm motion at the current time instant, the predicted motion using RNN is utilized as the measurement motion data in UKF.

UKF is a model-based filtering approach, that has commonly been implemented to track and predict human motion, e.g., [33]–[36]. Compared to the traditional UKF-based prediction studies obtaining the real-time measurement data by sensors, the proposed method utilizes neural networks to estimate the measurement data for a certain future horizon, and thus is beneficial for long-term motion prediction. Furthermore, different from [35], [36], the motion equations of the proposed method are derived using the Lagrangian approach, which directly connects the muscle force and human motion. In general, the main contributions of this work are summarized as follows. (1) An intrinsic connection between the future motion and the future muscle forces is explicitly established based on a Lagrangian-mechanics-based arm dynamic model. (2) A novel dynamic model informed prediction method is proposed. By incorporating the predicted muscle forces, UKF is capable of further improving the preliminary arm motion prediction obtained from the traditional RNN. (3) The effectiveness of the proposed dynamic model informed prediction method has been experimentally validated using three motion datasets.

The remainder of this paper is organized as follows. Section 2 presents an overview of the proposed approach. Section 3 describes the arm motion prediction using skeleton vectors. Section 4 introduces the proposed dynamic model informed arm motion prediction method. Section 5 presents the validation of the proposed prediction methods based on three motion datasets. Section 6 concludes this paper.

II. Overview of the Proposed Dynamic Informed Prediction Approach

This section introduces the overview of the proposed novel prediction approach. This method aims to utilize the dynamic model including future muscle forces to predict the future arm motion. The needed notations and definitions are presented as follows. The arm skeleton pose is denoted as $S = [v^u; v^f] \in \mathbb{R}^6$, where $v^u \in \mathbb{R}^3$ and $v^f \in \mathbb{R}^3$ are unit bone-vectors representing the directions of the upper-arm and forearm respectively. The observed skeleton pose sequence is $\mathbf{S} = [S_{-N+1}, \ldots, S_0] \in \mathbb{R}^{6\times N}$ and the predicted pose sequence is $\mathbf{S}^* = [S_1^*, \ldots, S_m^*, \ldots, S_M^*] \in \mathbb{R}^{6\times M}$, where m is the index of the prediction, and N and M are the time steps of the observation and prediction respectively. The observed and predicted muscle force sequences are denoted as $\mathbf{F}$ and

$\mathbf{F}^*$, respectively. $LSTM^S$ and $LSTM^F$ indicate two well-trained RNNs with LSTM structure, which are used to predict skeleton pose sequence $\mathbf{S}^*$ and force sequence $\mathbf{F}^*$. In addition, $\mathbf{x}$ stands for the predicted arm motion sequence after UKF.

Fig. 1 illustrates the overview of the proposed prediction approach. Firstly, captured motions are converted to the observed motion sequence $\mathbf{S}$, and $\mathbf{S}$ is utilized to obtain the predicted motion sequence $\mathbf{S}^*$, which is regarded as the measurement motion data in UKF. Secondly, to calculate the observed force sequence $\mathbf{F}$, the arm dynamic model is developed using Lagrangian mechanics and the inverse dynamics of the arm is formulated.

Thirdly, the arm dynamic model is converted to the motion transition and measurement models. Embracing the predicted force sequence $\mathbf{F}^*$ allows the motion transition model to calculate the future motion of arm joints. Eventually, UKF is leveraged to obtain the final predicted arm motion sequence $\mathbf{x}$.

III. Arm Motion Prediction Using Skeleton Vectors

This section presents how the human arm motion sequence $\mathbf{S}^*$ is predicted using skeleton vectors. We choose recurrent neural network with LSTM to capture temporal dependences of human arm's motion [37] and predict the future arm motion effectively. More precisely, the gates of LSTM assist the network to learn arm motion's characteristics, and the well-trained network utilizes the observed skeleton poses $\mathbf{S}$ to predict the future skeleton poses $\mathbf{S}^*$.

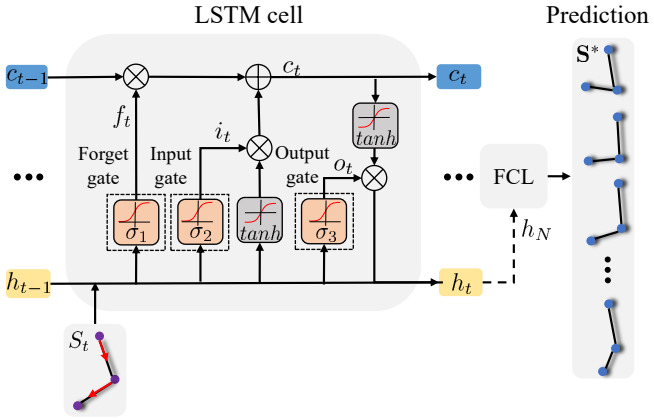


Fig. 2: The LSTM structure: purple dots indicate the observed arm joints, two red arrows are the directions of bones, FCL stands for the fully connected layer, the last hidden state h_N is the input of FCL, and blue dots represents the predicted arm joints.

We choose unit vectors of bones instead of 3D position data of arm joints to train a network. The advantage of such a choice is that by multiplying the unit vector with the corresponding bone length, the distance between two joints is always fixed when we re-construct the future arm motion. Fig. 2 presents the LSTM architecture, where c_{t-1} is the cell state memorizing the arm motion information from the previous iteration, h_{t-1} is the hidden state from the previous iteration, S_t is the current arm pose input, and σ and $tanh$ are the sigmoid and tanh activation functions. σ_1 is first applied as a forget gate to determine what motion characteristics from c_{t-1} need to be ignored using the following equation:

$$f_t = \sigma_1 \left(w_f \cdot [h_{t-1}, S_t] + b_f \right)$$

where f_t is the output of the forget gate, and w_f and b_f are the learning weights and bias of the forget gate layer. Next, σ_2 is applied as an input gate to determine what new motion characteristics need to be added into c_{t-1} using the following equation:

$$i_t = \sigma_2 \left(w_i \cdot [h_{t-1}, S_t] + b_i \right)$$

where i_t is the output of the input gate, and w_i and b_i are the learning weights and bias of the input gate layer. After that, the cell state c_{t-1} is updated to c_t using the following equation:

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh \left(w_c \cdot [h_{t-1}, S_t] + b_c \right)$$

where $\odot$ denotes the Hadamard product, and w_c and b_c are the learning weights and bias to update the cell state. Finally, σ_3 is applied as an output gate to output crucial motion characteristics using the following equations:

$$o_t = \sigma_3 \left(w_o \cdot [h_{t-1}, S_t] + b_o \right)$$
$$h_t = o_t \odot \tanh \left(c_t \right)$$

where o_t is the output of the output gate, and w_o and b_o are the learning weights and bias of the output gate layer. The last hidden state h_N is connected with a fully connected layer to provide the prediction $\mathbf{S}^*$. The prediction process is denoted as:

$$\mathbf{S}^* = LSTM^S(\mathbf{S}) \tag{1}$$

IV. Dynamic Model Informed Arm Motion Prediction

Although the skeleton vector-based LSTM effectively handles the stochastic nature of arm motion and successfully predicts arm motion, this purely data-driven model ignores the critical arm motion dynamics, which may cause motionless or occasional unrealistic predictions. Therefore, this section presents that we (1) implement Lagrangian mechanics to build the intrinsic connection between the arm motion and the joint's muscle force, and predict the future forces, and (2) incorporate the predicted force into the arm motion dynamic model and leverage UKF to obtain more accurate arm motion prediction.

A. Prediction of arm joints' muscle forces using dynamic model and RNN

This section describes the details on using Lagrangian mechanics to develop the arm dynamic model to compute the observed muscle forces acting on the shoulder and elbow joints, and applying LSTM to predict the future joint's muscle forces. Fig. 3 demonstrates the arm geometric model and two reference frames. To simplify the problem, we assume the shoulder is fixed on position O. The spherical shoulder frame I uses θ_1 and ϕ_1 to track the rotation of the shoulder joint based on the upper-arm motion. The spherical elbow frame E uses θ_2 and ϕ_2 to track the rotation of the elbow joint based on the forearm motion.

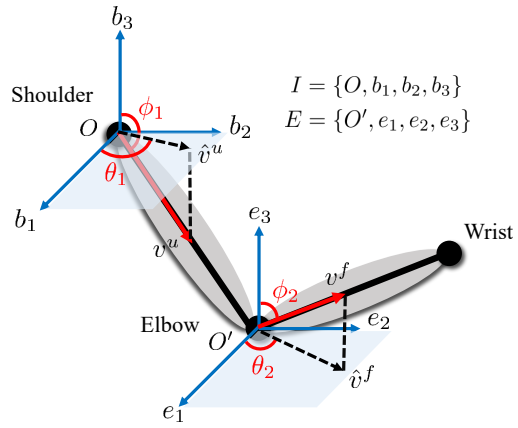


Fig. 3: Human arm geometric model and reference frame definition: the black dots represents joints, ϕ_1 is the angle between b_3 and v^u, ϕ_2 is the angle between e_3 and v^f, θ_1 is the angle between b_1 and $\hat{v}^u$, where $\hat{v}^u$ is the projection vector from v^u to b_1-b_2 plane, and θ_2 is the angle between e_1 and $\hat{v}^f$, where $\hat{v}^f$ is the projection vector from v^f to e_1-e_2 plane.

The Lagrangian mechanics is applied to derive the dynamic equations of arm motions. The arm Lagrangian L is defined using the total arm kinetic energy T to minus the total arm potential energy U with the following representation:

$$L = T - U$$

the Euler-Lagrangian equation [38] based on the derived Lagrangian function describes the arm motion dynamics and is given by:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) - \frac{\partial L}{\partial q} = F \tag{2}$$

where $q = [\theta_1; \theta_2; \phi_1; \phi_2]$ are the generalized coordinates and the non-conserved muscle forces F are the only considered generalized forces in this study. Eq. (2) are the second order ordinary differential equations and can be represented as the inverse dynamics matrix form

$$F = D(q)\ddot{q} + C(q, \dot{q})\dot{q} + G(q) \tag{3}$$

where $D(q)$ denotes a 4×4 symmetric and positive definite inertia matrix, $C(q, \dot{q})$ is a 4×4 velocity coupling matrix, $G(q)$ represents a 4×1 gravitational force vector, and $\dot{D}(q) - 2C(q, \dot{q})$ is a 4×4 skew-symmetric matrix [39].

The intrinsic relationship between the arm's motion and the muscle forces acting on the shoulder and elbow joints has been established based on Eq. (3). The observed generalized coordinate sequence $\mathbf{q} = [q_{-N+1}, \ldots, q_0] \in \mathbb{R}^{4\times N}$ as well as its first and second derivatives are used to generate the observed muscle force sequence $\mathbf{F} = [F_{-N+1}, \ldots, F_0] \in \mathbb{R}^{4\times N}$. We apply recurrent neural network with LSTM to predict the future muscle force sequence $\mathbf{F}^* = [F_1^*, \ldots, F_m^*, \ldots, F_M^*] \in \mathbb{R}^{4\times M}$ using the following equation:

$$\mathbf{F}^* = LSTM^F(\mathbf{F}) \tag{4}$$

The dynamic model integrating $\mathbf{F}^*$ can be used to predict the arm motion, and the details are illustrated in the next subsection.

B. Arm motion prediction using UKF

This subsection introduces the details of using UKF to predict the arm motion. The direct dynamics of the arm is represented with the following equation:

$$\ddot{q} = D(q)^{-1}(F - C(q, \dot{q})\dot{q} - G(q)) \tag{5}$$

The state of the arm dynamic model is denoted as $x = [q; \dot{q}] \in \mathbb{R}^8$ and regarded as the temporal arm motion, the continuous nonlinear motion transition function based on Eq. (5) is represented with the following equation:

$$\dot{x} = g_c(x, F) \tag{6}$$

By plugging the future muscle forces into Eq. (5), the future arm dynamics is obtained and used to estimate the future arm motion. Furthermore, Eq. (6) is converted to the following discrete-time motion transition function:

$$\begin{aligned} x_m &= x_{m-1} + g_c(x_{m-1}, F_{m-1}^*)T_s + \delta_m \\ &= g(x_{m-1}, F_{m-1}^*) + \delta_m \end{aligned} \tag{7}$$

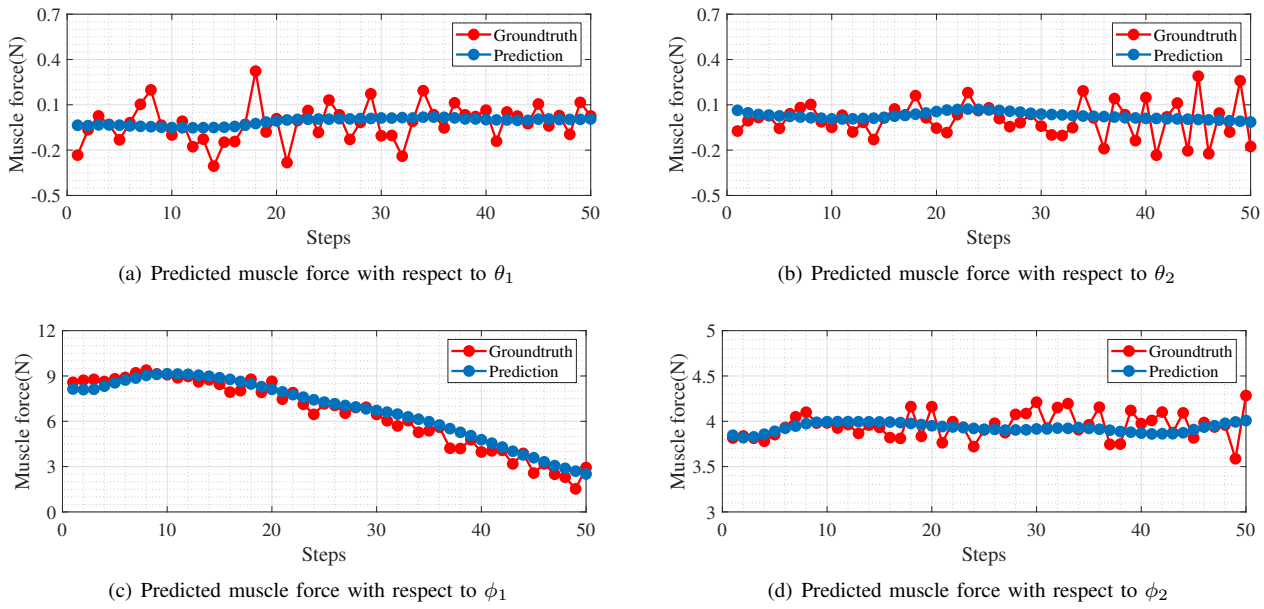
where $\delta_m \sim (0, Q_m)$ is process noise with Gaussian distribution. The motion measurement model is defined as the following equation:

$$y_m = h(x_m) + \zeta_m$$

where h is a mapping function from estimation space to measurement space, and $\zeta_m \sim (0, R_m)$ is measurement noise with Gaussian distribution.

Kalman filter has been widely applied to estimate the state of a system based on Gaussian distribution. However, the traditional Kalman filter is difficult to estimate the state of non-linear models. For example, to predict the arm motion for the next M time steps, the arm motion of the M time steps are assumed to be in the form of Gaussian distribution. However, the temporal arm motion x_m can not remain Gaussian since x_{m-1} passed the nonlinear function $g(\cdot)$. To deal with non-linear models, existing works, e.g., [40]–[42] leveraged extended Kalman filter (EKF) to first linearize the dynamic model, and then approximate the state distribution based on the linearized model. Despite the extended Kalman filter is capable of estimating the mean and covariance in many cases, the variables only propagate the first-order or second-order approximation of the true non-linear model, which brings large errors and is computational expensive in the estimation of highly non-linear systems [43].

To better estimate the arm motion, instead of propagating the linearized arm dynamic model like the way in extended Kalman filter, UKF applies the unscented transformation method [44] to enable the variables to propagate the true arm dynamic model, and approximate a new Gaussian. Assume the mean and covariance based on all estimated arm motions from the previous iteration are μ_{m-1} and P_{m-1}^x. And $2L+1$ sigma point vectors X_{m-1}^i is generated to construct a sigma matrix $\mathcal{X}_{m-1}$ representing the current whole distribution, where L is the number of features describing the temporal arm motion and $i = 0, 1, \ldots, 2L$. X_{m-1}^i is developed use the following equations [45]:



(a) Predicted muscle force with respect to θ_1

(b) Predicted muscle force with respect to θ_2

(c) Predicted muscle force with respect to ϕ_1

(d) Predicted muscle force with respect to ϕ_2

Fig. 4: Predicted muscle forces in motion A: (a) and (c) are the muscle forces acting on the shoulder joint, and (b) and (d) are the muscle force acting on the elbow joint.

$$
\begin{aligned}
X^0_{m-1} &= \mu_{m-1} \\
X^i_{m-1} &= \mu_{m-1} + \left(\sqrt{(L+\lambda)P^x_{m-1}}\right)_i \quad i = 1, \ldots, L \\
X^i_{m-1} &= \mu_{m-1} - \left(\sqrt{(L+\lambda)P^x_{m-1}}\right)_i \quad i = L+1, \ldots, 2L
\end{aligned} \tag{8}
$$

where λ is a scaling parameter, and $\left(\sqrt{(L+\lambda)P^*_{m-1}}\right)_i$ is the i-th column of the matrix root. The mean and covariance for the approximated new Gaussian are:

$$
\begin{aligned}
u_m &= \sum_{i=0}^{2L} W_i g(X^i_{m-1}, F^*_{m-1}) \\
\hat{P}^x_m &= \sum_{i=0}^{2L} W_i \alpha \alpha^T + Q_m.
\end{aligned} \tag{9}
$$

where $\alpha = g(X^i_{m-1}, F^*_{m-1}) - u_m$, and $W_0 = \lambda/(n+\lambda)$ and $W_i = 1/[2(n+\lambda)]$ are assigned weights. **Algorithm 1** describes the details of the proposed dynamic model informed arm motion prediction. To calculate the Kalman gain K_m, the mean and covariance of the measurement are calculated using the similar way shown in step 5. Importantly, considering we cannot measure the future arm motion at the current time instant, we utilize the predicted motion sequence $\mathbf{S}^*$ based on Eq. (1) to substitute for the sensor-based measurement data in UKF. In addition, $\mathbf{S}^*$ is converted to $\mathbf{y}^* = [y^*_1, \ldots, y^*_m, \ldots, y^*_M] \in \mathbb{R}^{8\times M}$, where $y^*_m = [q^*; \dot{q}^*] \in \mathbb{R}^8$. Eventually, y^*_m integrating the updated Kalman gain K_m is applied to predict the arm motion x_m shown in step 9.

V. Experimental Tests and Results

This section describes the experimental validation of the proposed arm motion prediction methods, we (1) setup an experimental platform to collect the arm motion data used for training, (2) train $LSTM^S$ and $LSTM^F$ to predict the skeleton pose sequence and muscle force sequence separately, and (3) predict the arm motion based on the future dynamic model and compare the results of the pure data-driven prediction and the dynamic model informed prediction.

A. Data acquisition

Experimental tests have been conducted to collect the motion data. The experiment platform is illustrated in Fig. 7. The human worker is doing grasping tasks. The Vicon camera system captures the arm motion based on the attached markers at a frequency of $25Hz$ and provides the arm's trajectory. The motion of the human arm has three types which are shown in Fig. 8. In motion A, the human worker arm first moves forward to grasp screwdrivers in the front side, then moves back to the original position. In motion B, the procedure of the arm movement is similar with the procedure in motion A. Nevertheless, the screwdrivers' location is on the left side of the human worker, which requires more rotation of shoulder and elbow joints. Moreover, in motion C, the human worker first grasps a screwdriver located on the left side. Next, the screwdriver needs to be dropped into the screwdriver box located on the right side. Eventually, the human hand moves back to its original position. Thus, the trajectories of arm are totally different in three motions.

B. Network training

We collected 132 trajectories for arm motion A, 144 trajectories for arm motion B, and 153 trajectories for arm motion C. For trajectories of each motion type, 70% is used for training, 15% is used for validation, and the remaining is used for testing. All trajectories are converted to skeleton vectors and muscle forces, separately. What's more, the anthropometric

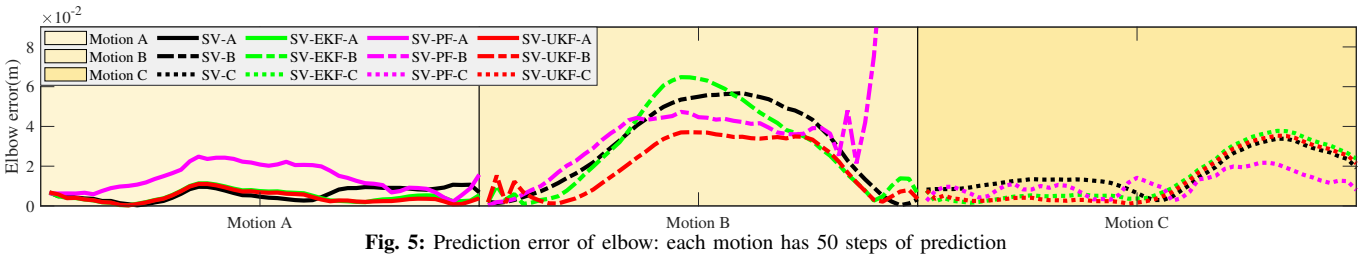


Fig. 5: Prediction error of elbow: each motion has 50 steps of prediction

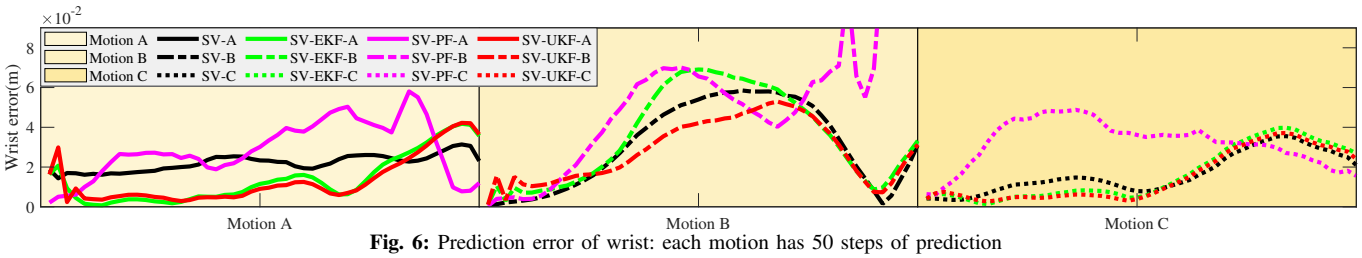


Fig. 6: Prediction error of wrist: each motion has 50 steps of prediction

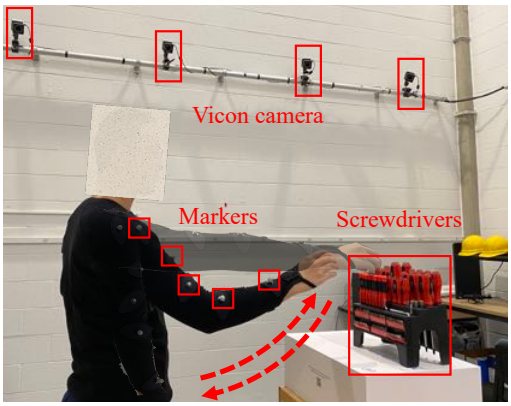


Fig. 7: Experiment platform

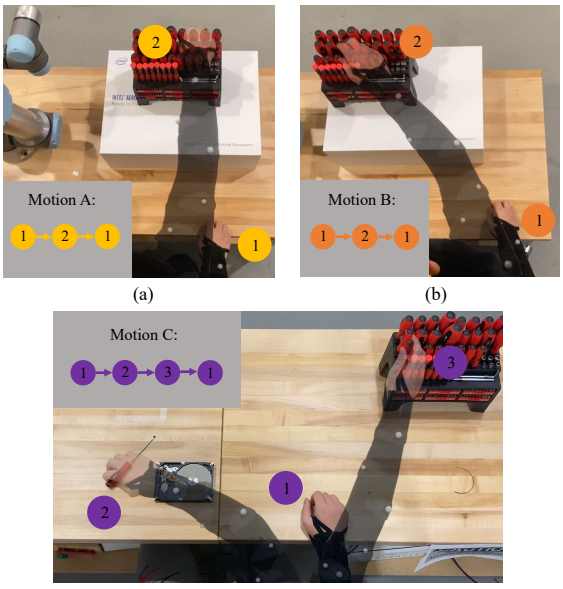


Fig. 8: Three types of arm motion: the numbers in the figure, e.g., 1, 2, and 3, represent the hand's position, screwdrivers are in front of the human worker in motion A, screwdrivers are on the left side of the human worker in motion B, and one screwdriver is on the left side and the remaining screwdrivers are on the right side in motion C.

data [46], e.g., bone length, masses and inertia moments of upper-arm and forearm, could be diverse based on the different human workers. We apply the common values of the anthropometric data to calculate the joint muscle forces based on the derived dynamic equations. We set the observation horizon N and prediction horizon M both to be 50, which indicates that the times of observation and prediction are both $2s$. We train the neural network for motion A, motion B, and motion C separately. For all well-trained networks, e.g., $LSTM^S$ and $LSTM^F$, the training loss function is the mean absolute error.

C. Prediction results and discussion

1) Muscle force prediction: The predicted and true muscle forces in motion A are demonstrated in Fig. 4. The four subplots correspond to the four generalized coordinates of q, separately. The predicted forces are obtained by a well-trained $LSTM^F$, and the ground-truth forces are calculated based on the inverse dynamics of the arm. Considering the Vicon system captures the arm motion at a high frequency, which introduces noise to the first and second derivatives of q, the calculated forces contain more noises than the predicted forces. Since we have the future muscle forces, a straightforward way to obtain the future arm motion would be solving Eq. (5) using integration. However, this approach integrates the future-force-based differential equation iteratively, and the calculated arm motion in the current iteration highly relies on the motion from the previous iteration. Although the predicted forces are very close to the true forces as shown in Fig. 4, the small error would accumulate throughout the whole integration, which leads to inaccurate or unrealistic predictions. Therefore, we leverage UKF to enable the prediction at each iteration to depend on the whole state distribution instead of only the prediction from the previous iteration, such that we could obtain a more accurate prediction.

Algorithm 1 Dynamic Model Informed Arm Motion Prediction

• Input: Observed arm motion $\mathbb{S}$ and $\mathbf{q}$
• Input: Well-trained networks $LSTM^S$ and $LSTM^F$
• Input: Arm dynamic model, Ts, R, and Q
1. Use $\mathbf{q}$ and its first and second derivatives to calculate $\mathbf{F}$ based on Eq. (3).
2. Obtain $\mathbf{F}^*$ based on Eq. (4).
for $m = 1, 2, \ldots, M$ **do**
3. Compute sigma points based on Eq. (8).
4. Calculate mean and covariance for the motion estimation based on Eq. (9).
5. Calculate mean and covariance for the measurement:

$$\hat{y}_m = \sum_{i=0}^{2L} W_i h(g(X^i_{m-1}, F^*_{m-1}))$$

$$P^y_m = \sum_{i=0}^{2L} W_i \beta\beta^T + R_m.$$

where $\beta = h(g(X^i_{m-1}, F^*_{m-1})) - \hat{y}_m$.
6. Calculate cross co-relation matrix P^{xy}_m between the estimation and measurement:

$$P^{xy}_m = \sum_{i=0}^{2L} W_i \alpha\beta^T$$

7. Update the Kalman gain K_m:

$$K_m = P^{xy}_m (P^y_m)^{-1}$$

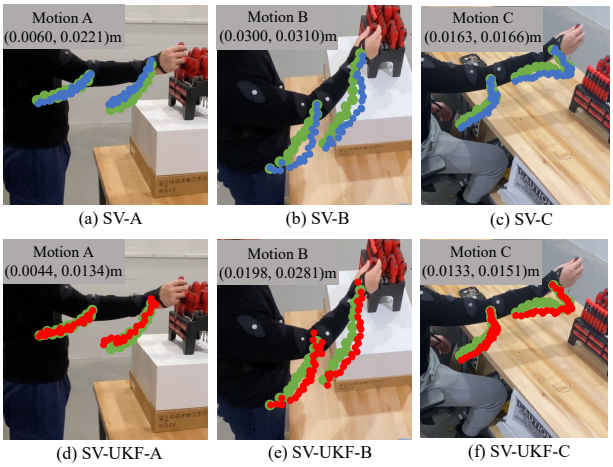
8. Obtain $\mathbf{S}^*$ based on Eq. (1), and convert $\mathbf{S}^*$ to $\mathbf{y}^*$.
9. Predict the arm motion and update the covariance:

$$x_m = x_{m-1} + K_m(y^*_m - \hat{y}_m)$$
$$P^x_m = \hat{P}^x_m - K_m P^y_m K^T_m$$

end
• Output: The arm motion sequence for the next M time steps $\mathbf{x} = [x_1, \ldots, x_m, \ldots, x_M]$.

2) Comparison: This paper also uses EKF and PF to predict arm motion as comparisons for UKF. EKF approximates the state distribution based on the linearized arm model instead of the true non-linear arm model like UKF and PF. In addition, the difference between UKF and PF is that instead of randomly choosing samples like PF, UKF employs the unscented transformation method to select sigma points to propagate the non-linear arm model.

To better illustrate our prediction results, the motion prediction method using skeleton vectors is denoted as SV, the dynamic model informed motion prediction methods using UKF, EKF, and PF with 4000 samples are denoted as SV-UKF, SV-EKF, and SV-PF, respectively. Furthermore, motion A, motion B, and motion C are represented as A, B, and C, respectively. The process noise is defined to be larger than the measurement noise considering the network prediction error and the anthropometric error, e.g., bone length, masses and inertia moments of upper-arm and forearm, both contribute to the process noise. The process noise for all motions is defined as 0.1. We tuned the measurement noises based on different motions. The measurement noises for motion A, B, and C are defined as (0.005,0.005,0.005,0.05), (0.05,0.003,0.06,0.003), and (0.02,0.005,0.01,0.005). For each kind of motion, SV-UKF, SV-EKF, and SV-PF all have the same measurement noise.



(a) SV-A (b) SV-B (c) SV-C
(d) SV-UKF-A (e) SV-UKF-B (f) SV-UKF-C

Fig. 9: Visualization of prediction: green, blue, and red trajectories indicate the ground-truth trajectory, the prediction using SV, and the prediction using SV-UKF respectively, and the average prediction error of the next 50 steps is labeled as (elbow error, wrist error)m

To qualify the motion prediction results, the prediction error plots of elbow and wrist at one-time instance are shown in Fig. 5 and Fig. 6, respectively. We use three kinds of lines to stand for three kinds of motions. The results from the proposed SV are represented with black color. The results from SV-UKF, SV-EKF, and SV-PF are represented using red, green, and pink colors. The comparison of the prediction errors indicates a few points. First, the SV-UKF's performance is close but slightly better than the SV-EKF's performance, especially in motion A and C. The other point is that, in motion C, despite SV-PF having the best performance in predicting the elbow's position, the prediction of the wrist's position is worst compared to other methods. In general, the predictions on the elbow and wrist using SV-UKF are more accurate and robust, thus, outperforms other methods.

To visualize the motion prediction results, the predicted arm trajectories and the average prediction errors are demonstrated in Fig. 9. The trajectories of the elbow and wrist in the next 50 steps are drawn manually. The ground-truth trajectories of the elbow and wrist are represented using green color, and the blue and red trajectories stand for the predicted trajectories using SV and SV-UKF, respectively. The average prediction errors of the elbow and wrist are shown in the top-left corner of each sub-figure. What's more, the percentage of the error reduction from SV to SV-UKF is shown in Table. I. By integrating the arm dynamic model, the average prediction error has been significantly reduced. Fig. 10 shows the effectiveness of SV-UKF in terms of the arm motion range. The prediction error over the motion range, most of the time, maintains below 5% in motion A. And for motion B and motion C, the maximum percentages are both below 15%.

Motion	Elbow Error Reduction	Wrist Error Reduction
A	26%	39%
B	34%	9%
C	18%	9%

TABLE I: Reduction percentage of average prediction error

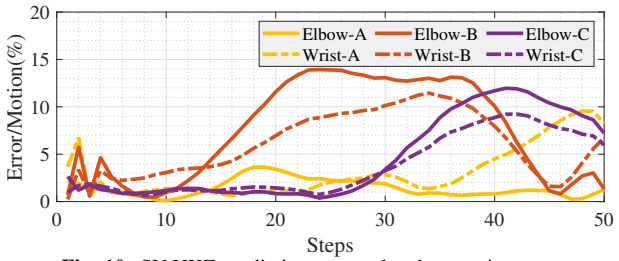


Fig. 10: SV-UKF prediction error related to motion range

VI. Conclusions and Future Work

This study has presented a novel arm dynamic model informed motion prediction method that aims to improve safety in HRC. We implement Lagrangian mechanics to develop the arm dynamic model that builds the connection between the arm motion and the corresponding muscle force. The past muscle forces are calculated based on the inverse dynamics of the arm, and the predicted muscle forces are obtained based on RNN. UKF is applied to handle the non-linear arm dynamic model and predict the future motion of the human arm. The proposed approach has been validated using three motion datasets, and the prediction results show that integrating the arm dynamic model indeed improves the prediction accuracy. Compared to the prediction error using the traditional RNN, the prediction error using the hybrid prediction method has been reduced significantly.

Considering the experimental tests only contain three kinds of motion, future studies can focus on making the proposed method to be generalized to different kinds of arm motions. Some generalization techniques, e.g., transfer learning, could be implemented to only use small sets of training data and networks that have promising prediction accuracy. Additional studies such as predicting the whole human body motion could also be achieved as long as the body dynamic model is constructed. Another area for future studies is to apply the proposed method based on a camera-based monitoring system since camera-based systems have been widely used in real industrial applications. The positional information of the human could be extracted and predicted from the captured RGB or RGB-depth images using some learning-based techniques, e.g., neural networks. Finally, the proposed method could also be beneficial for human-robot collaboration. For example, robot motion planning could take future human motion into account, such that human safety could be guaranteed when doing collaborative tasks.

References

[1] M.-L. Lee, S. Behdad, X. Liang, and M. Zheng, "Disassembly sequence planning considering human-robot collaboration," in *2020 American Control Conference (ACC)*, 2020, pp. 2438–2443.

[2] S. Sajedi, W. Liu, K. Eltouny, S. Behdad, M. Zheng, and X. Liang, "Uncertainty-assisted image-processing for human-robot close collaboration," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 4236–4243, 2022.

[3] M.-L. Lee, S. Behdad, X. Liang, and M. Zheng, "Task allocation and planning for product disassembly with human–robot collaboration," *Robotics and Computer-Integrated Manufacturing*, vol. 76, p. 102306, 2022.

[4] R. Fablet and M. J. Black, "Automatic detection and tracking of human motion with a view-based representation," in *European Conference on Computer Vision*. Springer, 2002, pp. 476–491.

[5] A. Bruce and G. Gordon, "Better motion prediction for people-tracking," in *Proc. of the Int. Conf. on Robotics & Automation (ICRA), Barcelona, Spain*, 2004.

[6] M. Kalakrishnan, S. Chitta, E. Theodorou, P. Pastor, and S. Schaal, "Stomp: Stochastic trajectory optimization for motion planning," in *2011 IEEE international conference on robotics and automation*. IEEE, 2011, pp. 4569–4574.

[7] J. Mainprice, R. Hayne, and D. Berenson, "Goal set inverse optimal control and iterative replanning for predicting human reaching motions in shared workspaces," *IEEE Transactions on Robotics*, vol. 32, no. 4, pp. 897–908, 2016.

[8] J. M. Wang, D. J. Fleet, and A. Hertzmann, "Gaussian process dynamical models for human motion," *IEEE transactions on pattern analysis and machine intelligence*, vol. 30, no. 2, pp. 283–298, 2007.

[9] G. W. Taylor, G. E. Hinton, and S. T. Roweis, "Modeling human motion using binary latent variables," in *Advances in neural information processing systems*, 2007, pp. 1345–1352.

[10] 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 *2011 IEEE International Conference on Robotics and Automation*. IEEE, 2011, pp. 5875–5880.

[11] C. Li, Z. Zhang, W. S. Lee, and G. H. Lee, "Convolutional sequence to sequence model for human dynamics," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 5226–5234.

[12] X. Liu, J. Yin, J. Liu, P. Ding, J. Liu, and H. Liub, "Trajectorycnn: a new spatio-temporal feature learning network for human motion prediction," *IEEE Transactions on Circuits and Systems for Video Technology*, 2020.

[13] Y. Cheng, W. Zhao, C. Liu, and M. Tomizuka, "Human motion prediction using semi-adaptable neural networks," in *2019 American Control Conference (ACC)*. IEEE, 2019, pp. 4884–4890.

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

[15] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*, 2014.

[16] H. Salehinejad, S. Sankar, J. Barfett, E. Colak, and S. Valaee, "Recent advances in recurrent neural networks," *arXiv preprint arXiv:1801.01078*, 2017.

[17] R. Liu and C. Liu, "Human motion prediction using adaptable recurrent neural networks and inverse kinematics," *IEEE Control Systems Letters*, vol. 5, no. 5, pp. 1651–1656, 2020.

[18] K. Fragkiadaki, S. Levine, P. Felsen, and J. Malik, "Recurrent network models for human dynamics," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 4346–4354.

[19] J. Martinez, M. J. Black, and J. Romero, "On human motion prediction using recurrent neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2891–2900.

[20] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," in *International conference on machine learning*. PMLR, 2013, pp. 1310–1318.

[21] Z. Liu, S. Wu, S. Jin, Q. Liu, S. Lu, R. Zimmermann, and L. Cheng, "Towards natural and accurate future motion prediction of humans and animals," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10 004–10 012.

[22] J. Hu, Z. Fan, J. Liao, and L. Liu, "Predicting long-term skeletal motions by a spatio-temporal hierarchical recurrent network," *arXiv preprint arXiv:1911.02404*, 2019.

[23] L.-Y. Gui, Y.-X. Wang, X. Liang, and J. M. Moura, "Adversarial geometry-aware human motion prediction," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 786–803.

[24] M. B. I. Reaz, M. S. Hussain, and F. Mohd-Yasin, "Techniques of emg signal analysis: detection, processing, classification and applications," *Biological procedures online*, vol. 8, no. 1, pp. 11–35, 2006.

[25] Q. Ding, A. Xiong, X. Zhao, and J. Han, "A novel emg-driven state space model for the estimation of continuous joint movements," in *2011 IEEE International Conference on Systems, Man, and Cybernetics*. IEEE, 2011, pp. 2891–2897.

[26] Q. Zhang, R. Hosoda, and G. Venture, "Human joint motion estimation for electromyography (emg)-based dynamic motion control," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2013, pp. 21–24.

[27] M. Pang, S. Guo, Q. Huang, H. Ishihara, and H. Hirata, "Electromyography-based quantitative representation method for upper-limb elbow joint angle in sagittal plane," *Journal of Medical and Biological Engineering*, vol. 35, no. 2, pp. 165–177, 2015.

[28] Y. H. Yin, Y. J. Fan, and L. D. Xu, "Emg and epp-integrated human–machine interface between the paralyzed and rehabilitation exoskeleton," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 4, pp. 542–549, 2012.

[29] J. Bae, K. Kong, and M. Tomizuka, "Real-time estimation of lower extremity joint torques in normal gait," *IFAC Proceedings Volumes*, vol. 42, no. 16, pp. 443–448, 2009.

[30] Z. Li, J. Sedlar, J. Carpentier, I. Laptev, N. Mansard, and J. Sivic, "Estimating 3d motion and forces of person-object interactions from monocular video," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 8640–8649.

[31] X. Lv, J. Chai, and S. Xia, "Data-driven inverse dynamics for human motion," *ACM Transactions on Graphics (TOG)*, vol. 35, no. 6, pp. 1–12, 2016.

[32] S. Cao and R. Nevatia, "Forecasting human pose and motion with multibody dynamic model," in *2015 IEEE Winter Conference on Applications of Computer Vision*. IEEE, 2015, pp. 191–198.

[33] H. Yang, X. Li, and X. Li, "Research on human motion data filtering based on unscented kalman filter algorithm," in *2021 33rd Chinese Control and Decision Conference (CCDC)*. IEEE, 2021, pp. 3269–3274.

[34] D. Lee, C. Liu, Y.-W. Liao, and J. K. Hedrick, "Parallel interacting multiple model-based human motion prediction for motion planning of companion robots," *IEEE Transactions on Automation Science and Engineering*, vol. 14, no. 1, pp. 52–61, 2016.

[35] Z.-Q. Zhang, W.-C. Wong, and J.-K. Wu, "Ubiquitous human upper-limb motion estimation using wearable sensors," *IEEE Transactions on Information technology in biomedicine*, vol. 15, no. 4, pp. 513–521, 2011.

[36] Z. Wang, S. Liu, and Y. Xu, "Human motion prediction based on hybrid motion model," in *2017 IEEE International Conference on Information and Automation (ICIA)*. IEEE, 2017, pp. 942–946.

[37] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "Lstm: A search space odyssey," *IEEE transactions on neural networks and learning systems*, vol. 28, no. 10, pp. 2222–2232, 2016.

[38] H. Goldstein, C. Poole, and J. Safko, "Classical mechanics," 2002.

[39] R. M. Murray, Z. Li, and S. S. Sastry, *A mathematical introduction to robotic manipulation*. CRC press, 2017.

[40] Z. Mynar, P. Vaclavek, and P. Blaha, "Synchronous reluctance motor parameter and state estimation using extended kalman filter and current derivative measurement," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 3, pp. 1972–1981, 2020.

[41] R. Hartley, M. Ghaffari, R. M. Eustice, and J. W. Grizzle, "Contact-aided invariant extended kalman filtering for robot state estimation," *The International Journal of Robotics Research*, vol. 39, no. 4, pp. 402–430, 2020.

[42] F. Lu, T. Gao, J. Huang, and X. Qiu, "A novel distributed extended kalman filter for aircraft engine gas-path health estimation with sensor fusion uncertainty," *Aerospace Science and Technology*, vol. 84, pp. 90–106, 2019.

[43] E. A. Wan and R. Van Der Merwe, "The unscented kalman filter for nonlinear estimation," in *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No. 00EX373)*. Ieee, 2000, pp. 153–158.

[44] S. J. Julier and J. K. Uhlmann, "New extension of the kalman filter to nonlinear systems," in *Signal processing, sensor fusion, and target recognition VI*, vol. 3068. International Society for Optics and Photonics, 1997, pp. 182–193.

[45] H. B. Khamseh, S. Ghorbani, and F. Janabi-Sharifi, "Unscented kalman filter state estimation for manipulating unmanned aerial vehicles," *Aerospace Science and Technology*, vol. 92, pp. 446–463, 2019.

[46] J. H. Challis and D. G. Kerwin, "Quantification of the uncertainties in resultant joint moments computed in a dynamic activity," *Journal of sports sciences*, vol. 14, no. 3, pp. 219–231, 1996.

Wansong Liu received the B.E. degree, in 2017, from China University of Mining and Technology, Xuzhou, China, and the M.S. degree in Mechanical and Aerospace Engineering, in 2020, from the University at Buffalo, NY, USA, where he is currently a Ph.D. student. His research interests include planning, learning, and control of robotic manipulators in collaboration with human.

Xiao Liang received the B.E. degree from Hunan University, Changsha, China, in 2010, and the M.S. and Ph.D. degrees from the University of California at Berkeley, Berkeley, CA, USA, in 2011 and 2016, respectively, all in civil engineering. He joined the Department of Civil, Structural and Environmental Engineering, University at Buffalo, NY, USA, in 2018. His research interests include health monitoring and autonomous inspection of infrastructure systems through advanced data analytics, model-based and machine learning.

Minghui Zheng received the B.E. and M.E. degrees, in 2008 and 2011 respectively, from Beihang University, Beijing, China, and the Ph.D. degree in Mechanical Engineering, in 2017, from University of California, Berkeley, USA. She joined University at Buffalo, NY, USA, in 2017, where she is currently an assistant professor in Mechanical and Aerospace Engineering. Her research interests include learning, planning, and control of multiple robotic systems such as collaborative manipulators for remanufacturing and drones for disaster resilience. She was the recipient of the NSF CAREER Award in 2021.