IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 1, 2020, accepted December 10, 2020, date of publication December 21, 2020,
date of current version December 31, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3045994

Data Driven Models for Human Motion Prediction in Human-Robot Collaboration

QINGHUA LI[1,3], ZHAO ZHANG[2,3], (Graduate Student Member, IEEE), YUE YOU[2,3], (Graduate Student Member, IEEE), YAQI MU[2,3], (Graduate Student Member, IEEE), AND CHAO FENG[1,3]

[1]Department of Physics, School of Electronic and Information Engineering, Qilu University of Technology (Shandong Academy of Sciences), Jinan 250353, China
[2]School of Electrical Engineering and Automation, Qilu University of Technology (Shandong Academy of Sciences), Jinan 250353, China
[3]Jinan Engineering Laboratory of Human-Machine Intelligent Cooperation, Jinan 250353, China

Corresponding author: Chao Feng (cfeng@qlu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61701270, and in part by the Young Doctor Cooperation Foundation of Qilu University of Technology (Shandong Academy of Sciences) under Grant 2017BSHZ008.

ABSTRACT To improve the safety and effectiveness of human-robot collaboration (HRC), the robot must plan a safe trajectory before the human movement is finished. Therefore, it is necessary to enable proactive robot behavior by making accurate intention prediction decisions early in a human motion. Furthermore, it is desirable to not only provide the long-term trajectory prediction of human motion but also characterize the uncertainty around it. In this paper, we present a human motion prediction framework to predict the motion trajectory of human arm in a reaching task. The proposed framework combines partial trajectory classification and human motion regression. By leveraging on the partial trajectory classification, our framework makes it possible to recognize the human action and to provide a trajectory prediction before the human movement is finished. The human motion regression can compensate the low accuracy of the representative trajectory through the fusion strategy. The proposed framework consists of two phases: online phase and offline phase. The offline phase aims to learn a regression model with optimized hyperparameters and a fusion strategy combining different prediction algorithms. In the online phase, based on the partial motion classification, the future reaching trajectory in a given time step is predicted by using a multi-step Gaussian process regression and representative trajectory. Experimental results show that our proposed framework achieved significant performance.

INDEX TERMS Human-robot collaboration, human motion prediction, Gaussian process regression, human action recognition, representative trajectory computation.

I. INTRODUCTION

With the continuous development of robotic technology, automatic pathological examination is also constantly improving. Pathological examination is to discuss the principle of human disease by observing cell morphology through a microscope. It requires making cells slides of diseased human tissues and staining the cells. Benefit from the reliability and high efficiency of robots, the time and cost of pathological examination are reduced. However, if the robot can safely interact with human partner in the same work volume, the efficiency of pathological examination can be further improved. This new working paradigm is particularly suitable for making tissue pathological slices. Because this kind of work is flexible and still has to be performed by humans, robots can reduce task completion time by transferring tools to humans or performing other auxiliary operations. The safe and successful execution of this scenario requires that the robot predicts human motions and subtly adjusts planned trajectories in real-time [1]. Recently, various motion planning methods have been proposed to improve the robot's ability to adjust the planned trajectories [2]–[4]. If human motion is predicted more accurately, it will help to adjust the robot's trajectory in advance and avoid collisions. Due to this requirement, predicting human motion using visual information plays a significant role in human-robot collaboration (HRC) [5].

Physics-based methods are widely used to predict human motion. This kind of method is based on a physical model of human motion and its parameters can be estimated both

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

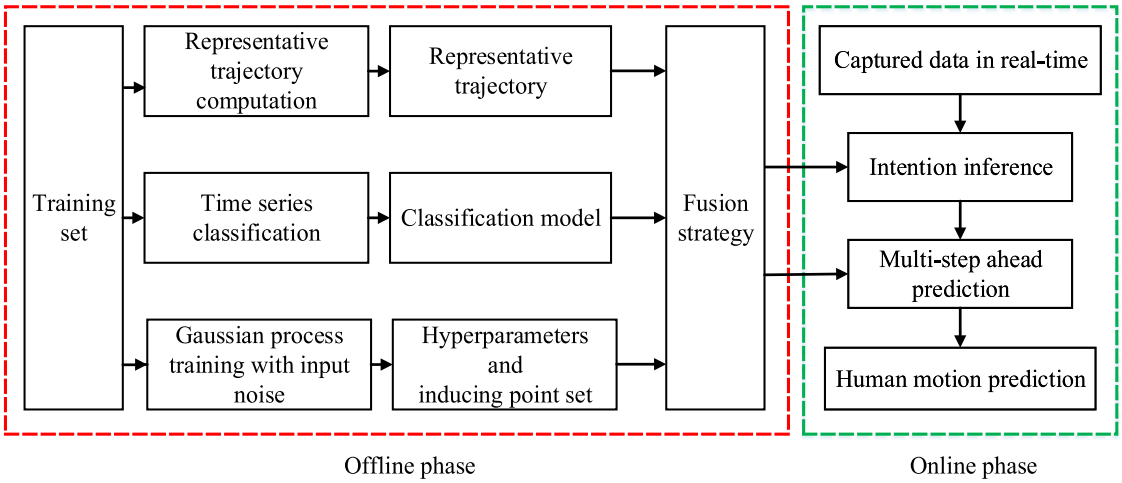


FIGURE 1. The structure block diagram of BP-HMT algorithm in offline and online phase.

online and offline by using system identification technologies. Examples include the use of constant acceleration model [6] for dynamic obstacle prediction based on Kalman filter, and hazard inference based on linear motion prediction of pedestrian [7]. Although these methods have achieved better performance, the physics-based method cannot provide the uncertainty of predicted human motion. Furthermore, minimum-jerk model is one of the most common deterministic methods, which assumes the speed profile of human motion fits the minimal jerk criterion. Under such assumption, human motion prediction algorithm based on the minimum-jerk model is integrated into the motion planning framework of HRC [8], [9]. Same to the aforementioned approaches, minimum-jerk model cannot provide confidence information for the results.

In addition, a lot of efforts have been placed in the area of human motion prediction using data-driven method. Data-driven methods utilize a statistical representation of human motion to predict human behavior. In the practical application of human motion prediction, there are challenges associated with gathering motions data using the current state of the art. First, the errors exist in the collected data impact the accuracy of human motion prediction due to the sensors or poor sampling [10]. Second, various uncertainties in terms of accurate representation of the environment will increase with the suddenly or abruptly change of human motion [11]. Compared with other methods, data-driven approach is advantageous for dealing with uncertainty. In [12], an online Gaussian process regression (GPR) is proposed to predict human hand trajectory. While this approach provides uncertainty information about the hand movement, it is still necessary to improve the accuracy of long-term prediction.

In this work, we present BP-HMT (Bayesian predictor for human motion trajectory), a framework for human motion prediction that synthesizes the classification and regression methods to predict the future human arm reaching motion. The BP-HMT algorithm includes a two-phase prediction. In the offline training phase, the motion-level anticipatory models are constructed using multiple demonstrations of human reaching motions. In the online prediction phase, a partial segment of the human arm trajectory is used for human intention inference through time series classification. Then, the human movement trajectory in a given time step is predicted using a multi-step recursive GPR and representative trajectory with the execution of the HRC task. This framework aims at solving long-term prediction problems, and its structure in the HRC system is shown in Fig. 1.

The major contribution of our work relies on the three points:

(1) We present an algorithm that consists of the inference of human intention and the prediction of future human motion. Compared with related methods, our algorithm can not only predict the future human position in a given time step, but also provide the uncertainty information of motion trajectory.

(2) The proposed method can deal with the noisy data in skeleton tracking in terms of human motion prediction.

(3) In the prediction of human movement trajectory, we propose a fusion strategy that combines the representative trajectory, classification model, and sparse GPR. The fusion strategy determines a trade-off between human motion regression and representative trajectory depending on their overall performance in multi-step ahead prediction, which improves the precision of human motion prediction.

This paper is organized as follows. In section II, the related works in the area of human motion prediction are discussed. Section III introduces the basic theory of Gaussian process regression. Section IV describes the component of the BP-HMT algorithm and a fusion strategy used in long-term human motion prediction. Simulation results and the algorithm performance are analyzed in Section V. Conclusions and future research directions are presented in Section VI.

II. RELATED WORK

In this section, we give a brief overview of prior work on human motion prediction in HRC.

In order to perform early motion recognition and predict the subsequent human motion, based on the target position at which the person arrives or strikes, Gaussian mixture model (GMM) was combined with Gaussian mixture regression to generate a representative human motion and estimate the task space area that the human will occupy [13]. Pérez-D'Arpino and Shah modified the GMM classification algorithm reported in [13] by adding a regularizer to prevent singularities [14]. For endowing GMM with better extrapolation performance, a task-parameterized formulation was studied in [15], which essentially models local trajectories and corresponding local patterns. Besides, an online unsupervised learning algorithm was presented to predict human motion, where the parameters of GMM are learned using the expectation-maximization (EM) algorithm [16]. Overall, most of the studies have focused on using GMM model to predict a goal position. We can not only compute the goal position that human is reaching toward, but also predict the trajectory of human motion at each time step.

Hidden Markov model (HMM) is another popular stochastic modeling technique for human motion [17]–[22]. Vasquez *et al.* used HMM to represent motion patterns, which can process observations incrementally by using the growing neural gas algorithm [19]. Ding *et al.* used the HMM to predict the regions in the workspace that are possibly occupied by humans, the regions can be served as constraints to avoid obstacles in robot motion planning [20]. Wang *et al.* proposed a novel approach for modeling the dynamics of human movements in the environment with a grid-based representation, a variant of the left-to-right HMM is used to directly model the moving tendency from the current cell to a neighbor cell [21]. Razin *et al.* used layered HMM with physiological data from arm muscles to achieve accurate human motion prediction [22]. Although the temporal information is addressed by the transition probabilities, the continuous distribution of trajectories cannot be represented completely with HMM.

In order to better predict the trajectory of human motion, it is necessary to use the dynamic time warping (DTW) [23] to calculate the similarity between different trajectories. Zhou and De la Torre developed generalized time warping (GTW), a technique for temporally aligning multiple multi-modal sequences (such as video, motion capture, and accelerometer data) from multiple subjects performing similar activities [24]. Although DTW is suitable for offline processing of data, the problem of online partial trajectory alignment is challenging. Pérez-D'Arpino and Shah used online DTW to calculate the normative length of each human activity to alleviate the problems caused by time deviation [14]. Furthermore, trajectory overlap and temporary stops can degrade the performance of existing alignment techniques. To address this problem, Lasota and Shah presented a Bayesian estimator that outputs the distributions of possible corresponding points based on observed partial trajectory data [25]. The main idea of these studies is to calculate the similarity of different trajectories by using DTW method. But the error between the prediction and ground truth value has increased based on partial trajectory. We use the corresponding point in the representative trajectory as predicted human position, which not only improves the accuracy of the prediction process, but also provides the uncertainty information of the predicted value.

In the process of human motion prediction, the robot must quickly infer the intention and future position of other human partners in HRC [26]–[33]. In order to predict human intentions, Ferrer and Sanfeliu denoted a complete probabilistic framework that consists of prediction algorithm, behavior estimator, and intentionality predictor [31]. A best destination of human reaching toward is estimated by intentionality predictor, which can greatly enhance the performance of the prediction algorithm. Quintero *et al.* proposed a balanced Gaussian process dynamical model (GPDM) to predict future pedestrian paths, poses, and intentions, which learns body motion dynamics of walking and stopping in a compact low-dimensional latent space [32]. Rehder *et al.* used a single artificial neural network (ANN) to solve the problem of intention recognition and planning-based prediction [33]. Koppula and Saxena proposed an anticipatory temporal conditional random field (ATCRF), which learns human intentions from a dataset with annotated object affordances (the functionality of the object), human activities, and human motion trajectories [34]. These studies mainly focus on the prediction of human intention and behaviors, which are not suitable for the prediction of long-term trajectory.

Given some demonstrations or observations, the inverse optimal control (IOC) algorithm is used to find the cost or reward function in the trajectory optimization problem. Berret *et al.* presented an automated IOC method to simultaneously examine several existing costs and any linear combination of them to automatically find the most suitable cost combination for human motion [35]. Mainprice *et al.* used the IOC algorithm to learn the cost function from the demonstration trajectory of the collaborative assembly task, the human motion from the current configuration to the target area is predicted by iteratively replanning the predicted trajectory using learned cost function [36]. Oguz *et al.* proposed a framework that combines the IOC method with the probabilistic motion primitive formulation, which can learn the motor variability as well as the interpersonal variance at the same time [37]. Besides, Ben Amor *et al.* extended the concept of imitation learning to human-robot interaction scenarios and introduced a new representation called interaction primitives [38]. This method predicts human intentions based on partial observations to adapt and correlate robot movements. Maeda *et al.* also used imitation learning to construct a mixture model of human–robot interaction primitive, which allows for both action recognition and human–robot movement coordination [39]. In [40], a hybrid motion prediction method that combines a minimum-jerk model and a dynamic movement primitives system is used to predict human

motion. Oguz *et al.* proposed a supervised learning framework to imitate close proximity dyadic interaction movement behavior and used recurrent neural networks (RNNs) to learn generalized policies [41]. Kratzer *et al.* proposed a human motion prediction framework, which uses a RNN to encode short-term dynamics and accounts for environmental constraints by gradient-based trajectory optimization [42]. Although the above methods perform well in the corresponding human motion prediction scenarios, the adjustment of model parameters is complicated.

Various effective fusion algorithms proposed in recent years have been applied to the field of human motion tracking and recognition. Glonek and Wojciechowski presented a data fusion method for human tracking, which can more accurately fuse the limb orientation data from two devices and compensate their imprecisions [43]. McIlwraith *et al.* proposed a framework to describe posture evolution by fusing the environment and wearable sensing patterns [44]. Bu *et al.* presented a hybrid motion classification framework that uses product rules to combine the probabilities of candidate motions obtained from the Bayesian network task model and the neural network classifier [45]. Ravichandar *et al.* proposed a gaze-based multi-model intention estimator to infer the target position of human reaching motions [46]. Pasciuto *et al.* introduced a hybrid motion prediction framework that relies on a reference captured motion database and combines motion control laws with a knowledge-based method, which is similar to the actual execution of the reference motion [47]. A multiple-predictor system that combines velocity-based position projection, time series classification, and sequence prediction is introduced in [48], which selects the optimal predictive output by a polynomial weighting algorithm. Xiang *et al.* proposed a hybrid predictive dynamics method to simulate human motion [49]. This method predicts human motion by combining optimization-based inverse kinematics, interpolation, and joint contour constraints. Cui *et al.* introduced a probabilistic fusion approach that combines low-dimensional and high-dimensional states for human motion tracking [50]. The difference between our work and the above work is that we consider the spatiotemporal characteristics of all offline trajectories to obtain representative trajectories, and combine multi-step ahead GPR to provide the uncertainty of all predicted values. The proposed fusion strategy provides long-term multi-step prediction results by weighing the prediction performance of different algorithms in all offline trajectories.

III. GAUSSIAN PROCESS REGRESSION

Gaussian process regression has been widely applied in the research of human motion prediction. It is a powerful and effective method to process nonlinear regression problems.

In Gaussian process regression [51], [52], a noise-free latent function $f(x)$ is approximated through a training data set $X \in R^{N\times D}$, $Y \in R^{N\times 1}$. Such that the relationship between the function $f(x_i)$ and the noisy output y_i are given by

$$y_i = f(x_i) + \varepsilon \tag{1}$$

where, $i = 1, \ldots, N$, $\{x_i, y_i\} \subset \{X, Y\}$, $\varepsilon \sim N(0, \sigma_n^2)$ is an independent identically distributed Gaussian white noise. To predict the value $f(X_*)$ at test vector X_*, the joint Gaussian distribution of the training output $f(X)$ and the predictive vector $f(X_*)$ (often shortened to f and f_*, respectively) is given as follows:

$$\begin{aligned}\begin{bmatrix} f^n \\ f_*^n \end{bmatrix} &\sim N\left(\begin{bmatrix} m(X) \\ m(X_*) \end{bmatrix}, \begin{bmatrix} k(X,X) & k(X,X_*) \\ k(X_*,X) & k(X_*,X_*) \end{bmatrix}\right) \\ &= N\left(\begin{bmatrix} m \\ m_* \end{bmatrix}, \begin{bmatrix} K & K_* \\ K_*^T & K_{**} \end{bmatrix}\right)\end{aligned} \tag{2}$$

where, $m(x)$ is the prior mean function for the Gaussian process and $k(x, x')$ is the prior covariance function. Note that we have introduced the shorthand notation at the second part of the equation. The superscript n denotes the number of training points, subscript $*$ denotes the parameter of the test set.

Many kinds of functions can be selected to construct the covariance matrix, and a common choice is the squared exponential covariance function:

$$k(x, x') = \lambda_f^2 \exp\left(-\frac{1}{2}(x - x')^T \Lambda^{-1}(x - x')\right) \tag{3}$$

where, λ_f^2 is a signal variance hyperparameter and Λ is a diagonal matrix of the squared lengthscale hyperparameter.

We can find the posterior distribution of both f and f_* given y using Bayes rule as:

$$\begin{aligned} p(f_*^n | f^n) &= \frac{p(f^n, f_*^n)}{p(f^n)} = N(\mu_*^n, \Sigma_{**}^n) \\ \mu_*^n &= m_* + K_*^{\mathrm{T}}(K + \Sigma_n)^{-1}(y - m) \\ \Sigma_{**}^n &= K_{**} - K_*^{\mathrm{T}}(K + \Sigma_n)^{-1}K_* \end{aligned} \tag{4}$$

where m and K are used to denote the properties of prior distributions, μ and Σ are used for posterior distributions.

IV. THE BP-HMT ALGORITHM FOR HUMAN MOTION PREDICTION

The BP-HMT algorithm combines representative trajectory computation, time series classification, and human motion regression. The first component of BP-HMT algorithm provides a set of spatially accurate representative trajectories of each task, which is achieved through calculating the average of all training trajectories at same durations. The next component analyzes the similarity between the current partial trajectory and each representative trajectory by DTW method [23]. Then, the representative trajectory that is most similar to the current partial trajectory is used to predict the position of human motion within a given time step. Furthermore, the third component of BP-HMT algorithm computes the prediction of joint movement trajectory by a multi-step ahead GPR, which improves the continuity of the predicted trajectory, reduces the prediction error, and calculates the uncertainty of human motion.

An important characteristic for data-driven human motion prediction algorithms is the full use of historical trajectory data. In this paper, we study the movement trajectories of the right hand. The trajectories of human joints are collected to build a training set D^T when human perform different tasks. Based on statistical analysis of tracked positions of human joints, the human motion toward several goal locations is also used to build a set of representative motions. Each trajectory is represented using T time frame of human motion, and k represents the instantaneous time.

A. REPRESENTATIVE TRAJECTORY COMPUTATION

The first step of BP-HMT algorithm is to define a set of representative trajectory Y^R, based on the training set D^T. The durations of different trajectory in the same task are simply normalized to a single nominal duration T^{nom}. Then, each trajectory is resampled by finding a cubic spline fit to the trajectory in a single nominal duration T^{nom}.

A probabilistic representation for each task as follows:

$$\begin{aligned}\mu_k &= \frac{1}{N_{dem}}\sum_{i=1}^{N_{dem}} Y_{i,k}^{nom} \\ \Sigma_k &= \frac{1}{N_{dem}-1}\sum_{i=1}^{N_{dem}} \left(Y_{i,k}^{nom}-\mu_{i,k}\right)\left(Y_{i,k}^{nom}-\mu_{i,k}\right)^{\mathrm{T}}\end{aligned} \quad (5)$$

where, μ is the mean vector and Σ is covariance vector of all the trajectories in the same task. Y^{nom} is the position vector of human motion with duration T^{nom} and k is time step, N_{dem} denotes N demonstrations of the same task.

As a result, the representative trajectory is obtained using the mean vector of each task. The representative trajectory of each task is denoted as a set of positions y_k^R sampled at time steps $k = 1, 2, \ldots, T^{nom}$. The movement trajectories of the right hand on the X-axis and its representative trajectory are shown in the Fig. 2.

B. TIME SERIES CLASSIFICATION

For the safety and efficiency of HRC, it is necessary to infer what humans are doing (e.g. take test tube from desktop) by the time series classification method. The first step of the time series classification is to search the representative trajectory of target motion class. Based on the observed on-going human motion, the similarity is measured among the observed partial trajectory and each of representative trajectories by using DTW method [23]. Then, the representative trajectory of target motion class is identified through the comparison of similarity. To predict the position $\hat{y}_{k+d_m}$ of human movement in a given time step d_m, a suitable representative point y_α^R that corresponds to the current observed position y_k should be identified. The representative point y_α^R is selected in a representative trajectory. The subscript α denotes the index of representative point that is closest to the current position, which can be obtained by calculating the minimum spatial

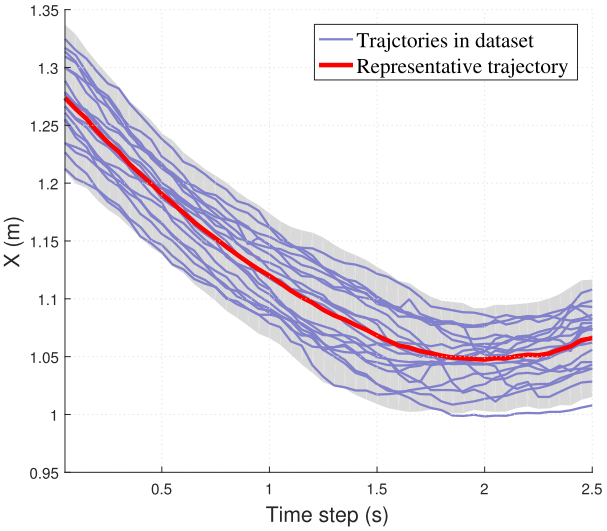


FIGURE 2. The blue solid line represents the trajectories resampled from the dataset, the red solid line represents the representative trajectory and the shaded area represents the variance per time step.

Euclidean distance:

$$\alpha = \underset{i\in[1,T^{nom}]}{argmin} \left\| y_i^R - y_k \right\| \quad (6)$$

After the representative point y_α^R is determined, the representative point will be used as a base point. Then, in the representative trajectory, the points within the range of multi-step ahead time d_m after the base point are used as predicted points. The predicted position $\hat{y}_{k+d_m}$ can be obtained as follows:

$$\hat{y}_{k+d_m} = y_{\alpha+d_m}^R \quad (7)$$

Due to the uncertainty of human motion, the error between the motion trajectory of each joint and the representative trajectory is inevitable. Although the representative trajectory can achieve better performance in the long-term prediction, there are still large errors in the initial stage of the action. Therefore, the human motion regression method is introduced to further improve the prediction accuracy. Based on the current observation point, the distribution of future points and their uncertainty can be obtained by combining human motion regression and representative trajectory.

C. HUMAN MOTION REGRESSION

Despite the excellent properties of GPR, there are still some obvious deficiencies. To reduce the number of parameters and the complexity of the computation, an active subset X_u is selected from training set. Then, we find the posterior distribution of the inducing outputs f_u at the corresponding inducing input points X_u.

$$\begin{aligned} p(f_u^n|f^n) &= N(\mu_u^n, \Sigma_u^n), \\ \mu_u^n &= m(X_u) + K_{uf}K_{ff}^{-1}(f - m(X)), \\ \Sigma_{uu}^n &= K_{uu} - K_{uf}K_{ff}^{-1}K_{fu} \end{aligned} \quad (8)$$

where, μ_u^n is the mean function and Σ_{uu}^n is the covariance of inducing point set.

Given a new input vector X_*, the joint distribution of inducing outputs f_u and predicted outputs f_* is expressed as:

$$\begin{aligned}\begin{bmatrix} f_u^n \\ f_*^n \end{bmatrix} &\sim N\left(\begin{bmatrix} \mu_u^n \\ \mu_*^n \end{bmatrix}, \begin{bmatrix} \Sigma_{uu}^n & \Sigma_{u*}^n \\ \Sigma_{*u}^n & \Sigma_{**}^n \end{bmatrix}\right) \\ \mu_*^n &= m(X_*) + K_{*u}K_{uu}^{-1}(\mu_u^n - m(X_u)), \\ \Sigma_{**}^n &= K_{**} - K_{*u}K_{uu}^{-1}(K_{uu} - \Sigma_{uu}^n)K_{uu}^{-1}K_{u*}. \end{aligned} \tag{9}$$

The sparse online noisy input Gaussian process (SONIG) [53] regression is an approach that tackles the incorporation of new noisy measurements in constant runtime and reduces the computation complexity by using online sparse GPR.

Therefore, when a new pair of input and output measurements (x_+, y_+) are available (the measurement is approximated using distribution $x_+ \sim N(\tilde{x}_+, \Sigma_{+x})$ and $y_+ \sim N(\tilde{f}_+, \Sigma_{+f})$), the distribution f_u is refined online by calculating the posterior distribution $p(f_u^{n+1}|\tilde{x}_+, \tilde{f}_+, f_u^n)$. The posterior distribution of the inducing input vector f_u can be calculated as follows:

$$\begin{aligned} f_u^{n+1} &\sim N(\mu_u^{n+1}, \Sigma_{uu}^{n+1}), \\ \mu_{u_i}^{n+1} &= \mu_{u_i}^{n+1}(\tilde{x}_+^{n+1}) + \frac{1}{2}tr\left(\frac{d^2\mu_{u_i}^{n+1}(\tilde{x}_+^{n+1})}{dx_+^2}\Sigma_{+x}^{n+1}\right), \\ \Sigma_{u_iu_j}^{n+1} &= \Sigma_{u_iu_j}^{n+1}(\tilde{x}_+^{n+1}) + \left(\frac{d\mu_{u_j}^{n+1}(\tilde{x}_+^{n+1})}{dx_+}\right)\Sigma_{+x}^{n+1}\left(\frac{d\mu_{u_i}^{n+1}(\tilde{x}_+^{n+1})}{dx_+}\right)^{\mathrm{T}} \\ &\quad + \frac{1}{2}tr\left(\left(\frac{d^2\Sigma_{u_iu_j}^{n+1}(\tilde{x}_+^{n+1})}{dx_+^2}\right)\Sigma_{+x}^{n+1}\right) \end{aligned} \tag{10}$$

where, $\mu_{u_i}^{n+1}$ and $\Sigma_{u_iu_j}^{n+1}$ denote the posterior distribution of individual elements. The superscript $n+1$ denotes the number of training points.

Additionally, our purpose is to predict a future trajectory of human motion. For the timely motion planning of robots, it is necessary that more information about human motion trends can be gathered. Hence, an iterative multi-step ahead prediction method is used for predicting future trajectory at time step k. However, according to the definition of the input and output in GPR, the prediction of outputs $\hat{y}_k, \ldots, \hat{y}_{k+d}$ depend on the given of measurements $\{x_i, y_i\}_{i=k}^{k+d}$. Therefore, when the y_k is estimated at current time step k, the distribution of $\hat{y}_{k+1}$ at time step k+1 is also obtained. Then, the distribution of $\hat{y}_{k+1}$ can be used as a virtual measurement to estimate the y_{k+1}. This progress is repeated until y_{k+d} is estimated. Furthermore, to reduce computation complexity, it is important that the inducing input points can be adjusted while the algorithm is running. It is possible to add a new measurement to the inducing point set when the new measurement is far from each of the inducing points. It is also possible to remove inducing input points between the current point and the closest inducing point to reduce computation complexity. We check if x_k is already close to any existing inducing input point, the normalized squared distance $dist(x_k, x_{u_i})$ can be calculated as follows:

$$dist(x_k, x_{u_i}) = (x_k - x_{u_i})^{\mathrm{T}}(x_k - x_{u_i}) \tag{11}$$

where, x_k is the current input point and x_{u_i} is the inducing input point.

If the normalized squared distance is below a given threshold, it means x_k close to the inducing input point x_{u_i}. Then x_{u_i}will be removed and x_k can be added to inducing point set X_u.

D. THE FUSION STRATEGY OF BP-HMT ALGORITHM

In the previous section, the basic structure of the BP-HMT algorithm is introduced. In this section, the fusion strategy and the two phases of BP-HMT algorithm are further described. Specifically, the two phases include offline phase and online phase. The basic prediction model is established in the offline phase, and the movement trajectory of right hand is predicted in the online phase.

In the offline phase of BP-HMT algorithm, each trajectory in the dataset is trained by noisy input Gaussian process (NIGP) [54]. The NIGP method is used to tune the hyperparameters h_{NIGP} by maximizing the marginal likelihood based on human motion trajectory. Then, the average value of hyperparameters and representative trajectories are indexed with label vectors in the database.

In order to obtain better prediction performance, we propose a fusion strategy that combines the prediction results of the time series classification and the human motion regression. In the offline phase, different methods are used to predict all motion trajectories in the training set. The prediction errors are calculated to evaluate the performance of each method at different stages. Then, a suitable prediction method can be selected after a switching time point k_s is given. The time point k_s can be calculated as follows:

$$\begin{aligned} E_k^{tsc} &= \left\| y_k - \hat{y}_k^{tsc} \right\| \\ E_k^{hmr} &= \left\| y_k - \hat{y}_k^{hmr} \right\| \\ k_s &= \underset{k\in[1,d_m]}{argmin} \sum_{i=1}^{N_{dem}} \left(E_{i,k}^{tsc} - E_{i,k}^{hmr}\right)^2 \end{aligned} \tag{12}$$

where, y_k is the ground truth value, $\hat{y}_k^{tsc}$ and $\hat{y}_k^{hmr}$ are predicted values calculated by time series classification and human motion regression, respectively. E_k^{tsc} and E_k^{hmr} represent the Euclidean distance between the ground truth and predicted value in the xyz spatial coordinate system. k_s is an algorithm switch point. d_m is the step length of multi-step ahead time.

In the online phase, the measurement data of the motion trajectory is collected for human action recognition and human motion regression. First, the DTW method is used to calculate the similarity between the current stage and the representative trajectory of each task. The label of the most similar task is used to obtain the representative trajectory y^R, hyperparameters h_{NIGP} and switching time step k_s of the corresponding task in the database. Then, the multi-step ahead regression and representative trajectory are combined

to calculate the predicted value. From different prediction methods, the predicted human motion and its uncertainty can be obtained. The online phase of the BP-HMT algorithm is summarized in Algorithm 1.

Algorithm 1 BP-HMT Algorithm for Human Motion Prediction in Online Phase

Input: The measurement point vector:
$(x_1, y_1), (x_2, y_2), \ldots, (x_{k-1}, y_{k-1})$
Output: Predicted value: $\hat{y}^o_{k+1}, \ldots, \hat{y}^o_{k+d_m}$
Initialization: $k_{Threshold}$, $dist_{Threshold}$
while True **do**
  Wait for a new measurement point: (x_k, y_k)
  if $k < k_{Threshold}$ **then**
    $label_t = \underset{t=1,\ldots,n}{argmin}\, dist_{DTW}(\{y_i\}_{i=1}^{k}, \{y_i^{R,t}\}_{i=1}^{k})$;
    (y^R, h_{NIGP})= $database(label_t)$;
    Obtain the switch time step k_s from database.
  else
    Calculate α using Eq. (6);
    Calculate $\mu^k_{u_i}$, $\Sigma^k_{u_iu_j}$ using Eq. (10);
    Calculate μ^k_*, Σ^k_{**} using Eq. (9);
    for $d = 1$ **to** d_m **do**
      Calculate $\hat{y}^{tsc}_{k+d}$ using Eq. (7);
      $X_u = X_u \bigcup x_{k+d}$;
      Calculate $\mu^{k+d}_{u_i}$, $\Sigma^{k+d}_{u_iu_j}$ using Eq. (10);
      Calculate μ^{k+d}_*, Σ^{k+d}_{**} using Eq. (9);
      Obtain $\hat{y}^{hmr}_{k+d}$ from μ^{k+d}_*.
      if $dist(x_{k+d}, x_{u_i}) \leq dist_{Threshold}$ **then**
        $X_u = X_u \backslash \{x_{u_i}\}$;
      end if
    end for
    The predicted output:
    $\{\hat{y}^o_i\}_{i=k+1}^{k+d_m} = \{\hat{y}^{hmr}_i\}_{i=k+1}^{k+k_s} \bigcup \{\hat{y}^{tsc}_i\}_{i=k+k_{s+1}}^{k+d_m}$
  end if
  $k = k + 1$;
end while

V. EXPERIMENTAL

To evaluate the performance of BP-HMT algorithm, a tabletop manipulation task is conducted. The human motion from three participants is captured by a Kinect depth sensor operating at a 20 Hz frame rate, and only upper body joints are tracked for action recognition. The prediction performance of BP-HMT algorithm for reaching motion with different target positions is discussed. The initial state of human is a static pose in front of a table. Since we want to cover a wide range of reaching motions, each participant is required to reach eight targets from the starting position facing the table, as shown in Fig. 3. In the online prediction process, $k_{Threshold}$ is set as 10, and $dist_{Threshold}$ is set as 0.1. The number of inducing point of BP-HMT algorithm is updated online. All the experiments are performed on the MATLAB 2015 platform with a 1.5GHz Intel core i3 processor.

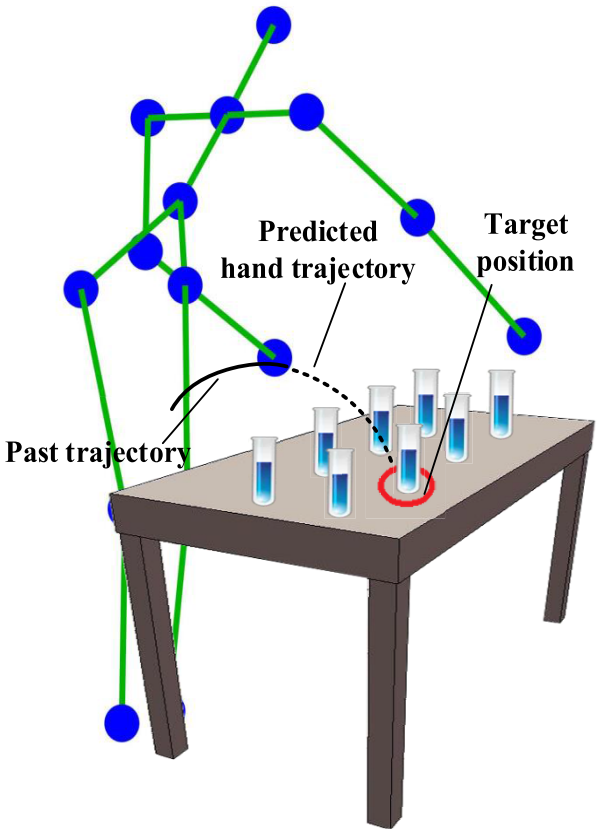


FIGURE 3. Prediction of hand movement trajectory in reaching tasks with different target positions. The test tube that marked with red circle is the one that human will grab. The future trajectory of hand movement (dash line) is predicted based on the past trajectory (solid line).

The SONIG-based multi-step ahead regression [12] is compared with the proposed BP-HMT algorithm in a general HRC environment. A target position can be inferred naturally when human arm reaching toward an object on the table. Then, the future human motion trajectory is generated recursively by SONIG algorithm using noisy input points in an online manner. However, compared with the SONIG algorithm, the BP-HMT algorithm can deliver reliable online human motion prediction even for a long prediction time. Therefore, it is necessary to compare the performance of BP-HMT and SONIG methods by further measuring the accuracy of multi-step ahead prediction.

Furthermore, the BP-HMT algorithm is also compared with the Bezier curve [55] and the minimum-jerk model [56]. Four position points are used to define the shape of the trajectory in the Bezier curve: action start point, action endpoint, and two control points. The start position of the trajectory is used as start point, the action endpoint and two control points are selected from the representative trajectories in the training set. Four parameters are also used to define the trajectory curve in the minimum-jerk model: the current position, the end position of the human movement, the current speed, and endpoint speed. The current position

and speed are obtained using real-time motion information of human joints. The position and speed of the end-point are selected from the representative trajectories in the training set.

In order to evaluate the performance of different methods, mean absolute error (MAE) and mean relative error (MRE) are selected as numerical criteria:

$$e_x(t) = x_d(t) - x_r(t),$$
$$e_y(t) = y_d(t) - y_r(t),$$
$$e_z(t) = z_d(t) - z_r(t); \tag{13}$$

$$E_{MAE} = \frac{1}{N}\sum_{t=1}^{N}\sqrt{(e_x(t))^2+(e_y(t))^2+(e_z(t))^2} \tag{14}$$

$$E_{MRE} = \frac{1}{N}\sum_{t=1}^{N}\sqrt{\left(\frac{e_x(t)}{x_r(t)}\right)^2+\left(\frac{e_y(t)}{y_r(t)}\right)^2+\left(\frac{e_z(t)}{z_r(t)}\right)^2} \tag{15}$$

where, in a xyz spatial coordinate system, $x_d(t)$, $y_d(t)$, and $z_d(t)$ are predicted values at time $k+d$. $x_r(t)$, $y_r(t)$, and $z_r(t)$ are ground truth values at time $k+d$.

A. HUMAN ACTION RECOGNITION

In this section, we evaluate the effectiveness of BP-HMT algorithm in observing different durations of human trajectories. We recorded eight sets of motion trajectories when human reaching toward an object at different positions on the table, and the scenario as shown in Fig. 3. The BP-HMT algorithm is used to calculate the accuracy of action recognition for each set of trajectories. A leave-one-out cross validation (LOOCV) is used as evaluation method. In each validation, one action is used as test set, others actions as training set. The recognition performance of the BP-HMT algorithm for the three sets of trajectories in the recorded actions is shown in Fig. 4.

The action recognition accuracy according to the partial trajectory labeled with corresponding action is shown in Fig. 4 (a). The red circle marks represent the case when the trajectory of the action 1 is used as test data. The blue square marks represent the case when the trajectory of the action 2 is the correct action, and the black cross marks represent the case when the action 3 is the correct action. The data measured from the motion capture system is used as a noiseless case. When 60% of the task is observed, the action 1 can be correctly recognized in all tests. Furthermore, the noisy data is also necessary to evaluate the performance of human action recognition. As shown in Fig. 4 (b), Gaussian noise with zero mean and standard deviation of 3 cm is added on the observed point. The observations are changed by the superimposed noise, which increases the possibility of false human motion recognition. Fig. 4 (c) shows the case that the Gaussian noise has a standard deviation of 6 cm. In this case, long-term observation is required to obtain higher recognition accuracy.

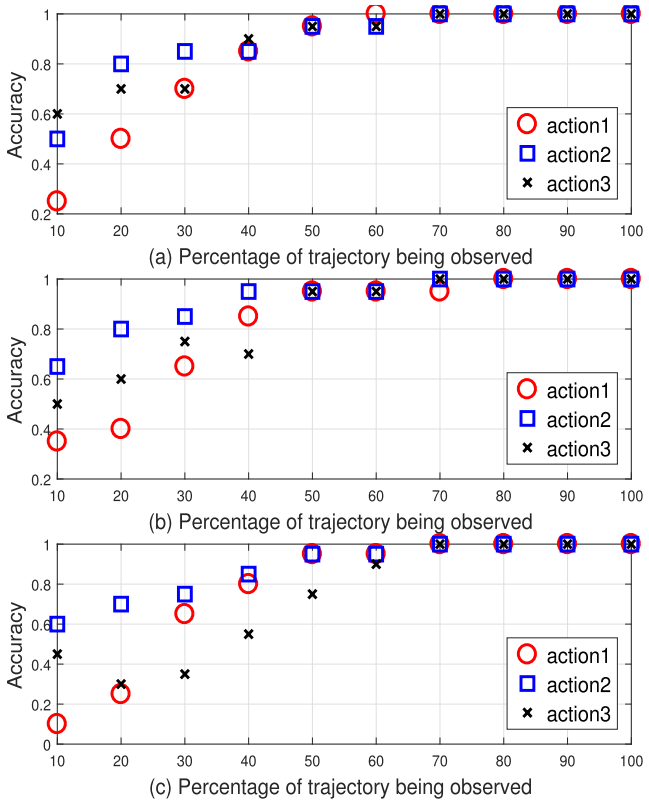


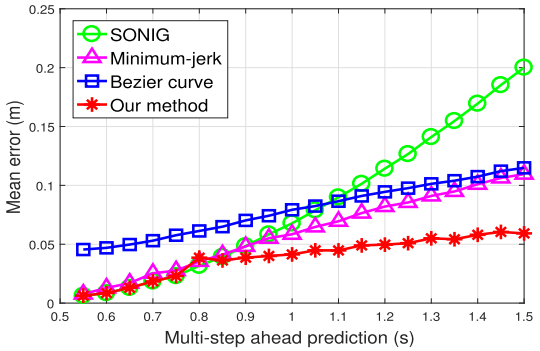
FIGURE 4. The accuracy of human action recognition with different amounts of noise. (a) Noiseless observation. (b) Observations added with zero-mean Gaussian noise and standard deviation 3 cm. (c) Observations added with zero-mean Gaussian noise and standard deviation 6 cm.

B. COMPARISON WITH RELATED WORKS

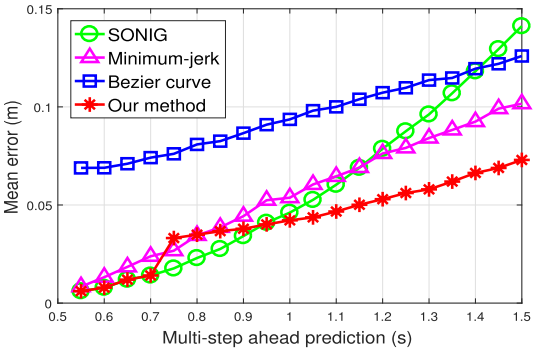
While we focused on human action recognition in the previous experiments, the human motion trajectory lasting 1s is predicted in this section. The motion trajectories of the right hand reaching different positions on the desktop are saved in different datasets, and each dataset records 20 hand motion trajectories. The prediction results of four methods are compared in eight trajectory sets. Furthermore, the range of multi-step ahead times considered is 0.5s to 1.5s, in increments of 0.05s. As shown in Fig. 5, the long-term prediction accuracy of the BP-HMT algorithm is higher than others, and the reasons for its performance are futher analyzed.

In order to analyze the experiment results, the prediction errors of all trajectories are computed for each of the four methods (i.e. the SONIG algorithm, Bezier curve, minimum-jerk model and our BP-HMT algorithm). These results are combined into separate vectors to calculate the mean prediction errors of each prediction method at corresponding time step.

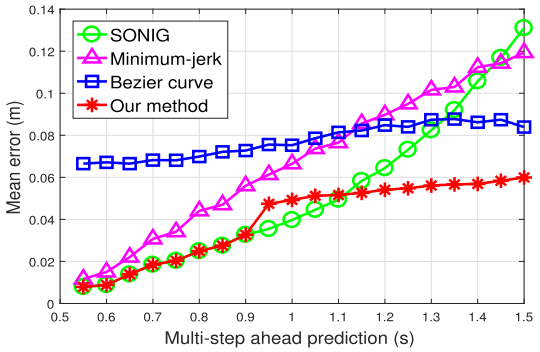
Compared with the SONIG and BP-HMT algorithm, the mean prediction error of the Bezier curve and minimum-jerk model at initial stage of multi-step ahead prediction (e.g. 0.55s to 0.7s in Fig. 5 (b), or 0.55s to 0.9s in Fig. 5 (c)) is higher. The reasons for this case are analyzed: the Bezier curve and minimum-jerk model rely



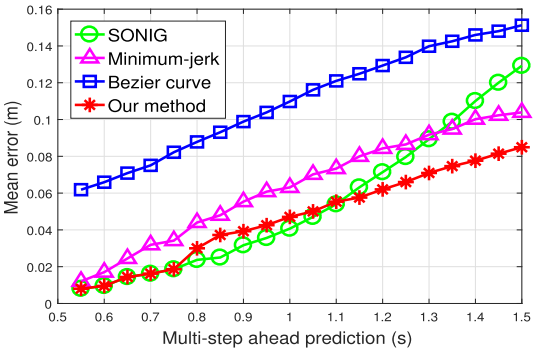
(a) task 1



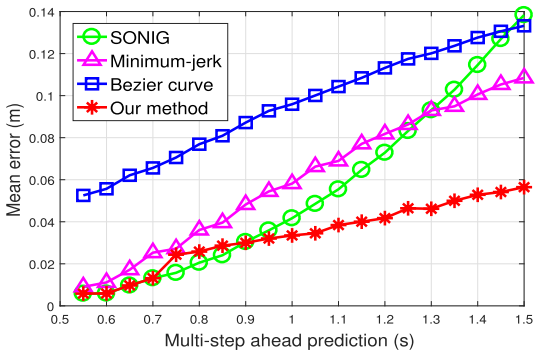
(b) task 2



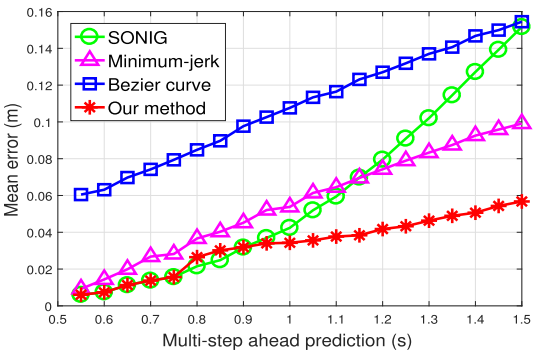
(c) task 3



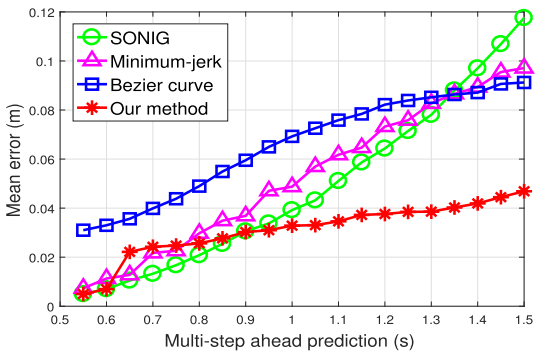
(d) task 4



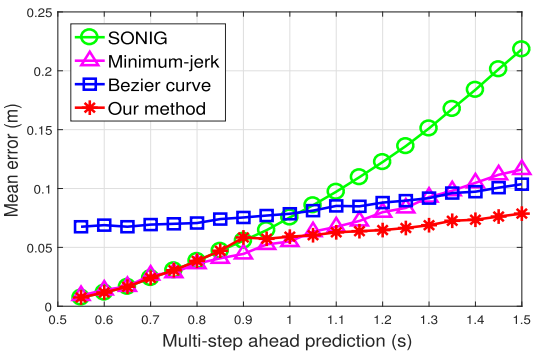
(e) task 5



(f) task 6



(g) task 7



(h) task 8

FIGURE 5. **The overall mean errors of multi-step ahead prediction. The BP-HMT algorithm is compared with other methods in eight test sets.**

TABLE 1. The overall MAE of human motion prediction in eight task test sets.

Method	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7	Task 8
SONIG [12]	0.0838	0.0585	0.0524	0.0543	0.0552	0.0599	0.0489	0.0923
Bezier curve [55]	0.0797	0.0954	0.0773	0.1101	0.0959	0.1085	0.0657	0.0819
Minimum-jerk [56]	0.0605	0.0565	0.0680	0.0639	0.0604	0.0567	0.0528	0.0607
Our method	0.0395	0.0421	0.0402	0.0472	0.0335	0.0332	0.0312	0.0519

TABLE 2. The overall MRE of human motion prediction in eight task test sets.

Method	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7	Task 8
SONIG [12]	1.3058	1.2067	0.1767	0.6208	1.1703	3.7789	0.0937	0.1482
Bezier curve [55]	1.5012	1.2784	0.2467	0.8737	1.7949	6.5562	0.1558	0.2747
Minimum-jerk [56]	1.4723	1.3851	0.2737	0.5823	1.2113	2.2370	0.1318	0.1772
Our method	0.7499	0.8163	0.1535	0.5175	0.6570	2.8058	0.0822	0.1713

on four parameters to fit the overall trajectory of hand motion. But the joints position or velocity can be affected by noise or sudden change of limb movement direction, which will change the parameters of the model and produce large prediction errors. Due to the representative trajectory is the statistical mean of multiple motions, the higher error is also produced using representative trajectory when the variance is large. In order to deal with this problem in the initial stage mentioned above, the iterative multi-step regression is combined with representative trajectory through the fusion strategy to reduce the prediction error. In addition, the iterative multi-step regression can also provide the calculation of uncertainty.

Among the long-term predictions of the four methods, the mean prediction error of the BP-HMT is lower. Although the Bezier curve and the minimum-jerk model also exhibited a trend toward decreased performance at multi-step ahead time, their mean error growth rate is lower than that of the SONIG method. This case is further analyzed as follows: Because the overall trajectory movement trend is predicted in the Bezier curve and the minimum-jerk model, the prediction accuracy of the model is limited by four parameters. On the contrary, since the long-term prediction of SONIG method is not constrained by the target position, it is conceivable that the prediction error continues to increase. In addition, the BP-HMT method uses the corresponding points on the representative trajectory as the output, so it has the smallest error in long-term trajectory prediction.

C. STATISTICAL ANALYSIS OF THE PREDICTION RESULTS

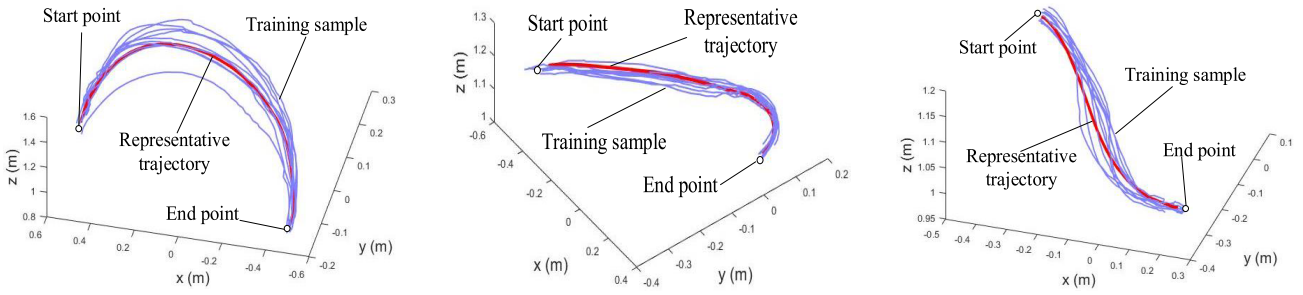
To analyze the performance of each method, the MAE of each trajectory in test set is computed. These results are combined into separate vectors to calculate the overall MAE in each task. As shown in Table 1, the accuracy of the BP-HMT algorithm in long-term prediction is higher than other algorithms in eight human reaching tasks. The MRE of each trajectory is also computed, and the overall MRE in each task is shown in Table 2.

The overall mean error (across all datasets) of the BP-HMT algorithm is 36.9%, 55.3%, and 33.5% lower than that of the SONIG, Bezier curve, and minimum-jerk model, respectively. As indicated by the results in Table 1, the BP-HMT nearly always outperformed the other three prediction methods. The prediction accuracy of BP-HMT is improved most obviously in the task 1, task 6 and task 5. Compared with SONIG, Bezier curve and minimum-jerk model, the prediction accuracy is improved by 52.9%, 69.4% and 44.5%. But in the task 3 and task 4, the prediction accuracy of the BP-HMT algorithm is only 13.1%, 36.6%, and 14.5% higher than that of the SONIG, Bezier curve, and minimum-jerk model, respectively.

Furthermore, the MRE is also used to evaluate the performance of the long-term prediction, as shown in Table 2. The MRE of the BP-HMT is 29.9%, 53.0%, and 20.3% lower than that of the SONIG, Bezier curve, and minimum-jerk model, respectively. The results of BP-HMT algorithm nearly always outperformed the other prediction methods, except for task 6 and task 8. The prediction accuracy of BP-HMT algorithm is improved most obviously in the task 1 and task 5. Compared with SONIG, Bezier curve and minimum-jerk model, the prediction accuracy is improved by 43.9%, 63.4% and 49.0%. But in the task 7, task 2 and task 4, the prediction accuracy of the BP-HMT algorithm is only 12.2%, 36.1%, and 11.1% higher than that of the SONIG, Bezier curve, and minimum-jerk model, respectively.

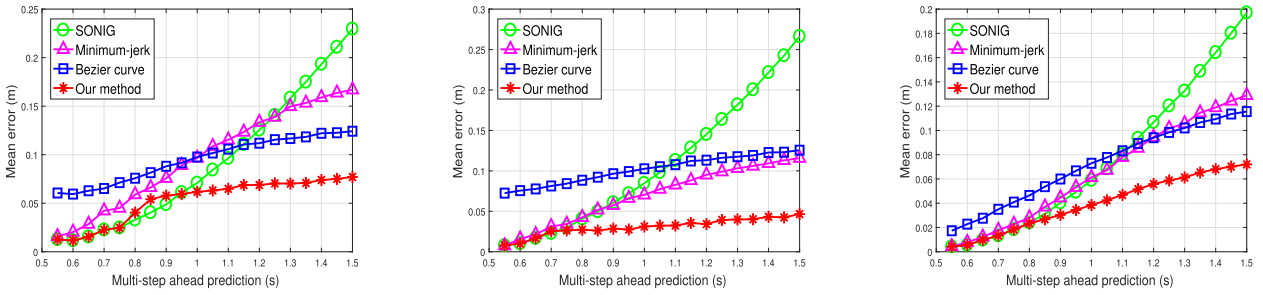
D. EXPERIMENTAL RESULTS OF TYPICAL TASKS

In the above work, our reaching task is to move the hand from the same initial position to different positions on the desktop. This reaching task is used to test the long-term prediction performance of different methods. However, the BP-HMT algorithm needs to be discussed with the typical trajectories of different start and end positions in general HRC scenario. Therefore, the hand movement trajectories in three typical tasks are collected to further evaluate the prediction performance of the BP-HMT algorithm. As shown in Fig. 6, three



(a) task 1 (b) task 2 (c) task 3

FIGURE 6. Human hand movement trajectories when performing different tasks. The three typical tasks are as follows: (a) take test tube from desktop, (b) reaching tube holder, and (c) transferring test tube. The blue solid lines are training samples while performing a task. The red solid line shows the representative trajectory of hand movement.



(a) task 1 (b) task 2 (c) task 3

FIGURE 7. The overall mean errors of multi-step ahead prediction. The BP-HMT algorithm is compared with other methods in three typical tasks.

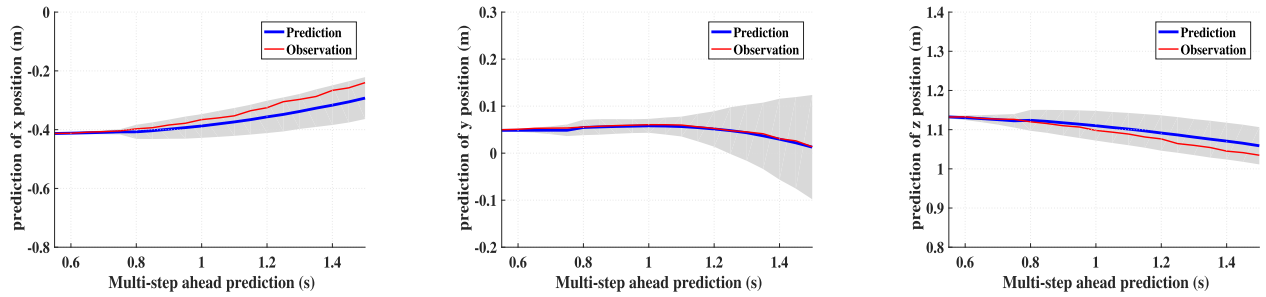


FIGURE 8. Prediction results of the BP-HMT algorithm. The blue solid line represents the prediction trajectory. The red solid line represents the observation. The shaded area represents the variance of x, y, z positions per time step in Cartesian space, respectively.

typical tasks are: take test tube from desktop, reaching tube holder, and transferring test tube.

The overall mean errors of the predicted trajectory at each time step are shown in Fig. 7. The same LOOCV procedure is used to evaluate the performance of long-term prediction, where one trajectory is used as the test set, the other trajectories are used as the training set. The overall mean errors are calculated with the error of each corresponding time step in all test sets. In Fig. 7 (a) to (c), the overall mean error of the proposed BP-HMT algorithm is lower than the other three methods (SONIG, minimum-jerk model, and Bezier curve). But for 0.8 s to 0.9 s in Fig. 7 (a), the prediction error of BP-HMT algorithm is higher than that of the SONIG algorithm. Similar to the previous analysis, although the prediction results are constrained in the long-term prediction using representative trajectories, the prediction results are still affected by the variance of the training set. An example of a predicted trajectory is shown in Fig. 8, which is defined by predicted values and corresponding variances in all dimensions. The spatial position of the observation at the corresponding time step is calculated by BP-HMT algorithm. With the increase of step size, the error also increases. But the BP-HMT algorithm can provide the variance of the predicted trajectory. Therefore, the safety in the HRC can be further improved by considering the uncertainty of human behavior.

VI. CONCLUSION AND FUTURE WORK

In this paper, we proposed an approach that consists of representative trajectory computation, time series classification, and human motion regression. In the offline phase of the proposed algorithm, a probabilistic model of human motion with optimized hyperparameters is trained and the representative trajectory of each task is computed. Based on the learned model and newly obtained measurements, the time series classification and sparse GPR are used for multi-step ahead prediction in the online phase. To improve the prediction accuracy of representative trajectory, we adopt a fusion strategy to obtain the optimized fusion point to combine the results of time series classification and human motion regression. The results show that with the combination of human motion classification and regression, our approach can extract and understand the human intention, which have better performance of long-term human motion prediction.

In future work, we will investigate the scalability of the approach. Currently, supervised learning with labeled action types is learned to recognize human action. More variety and size of captured data is processed and learned with the requirement of better predictive performance in general. To recognize some action and the human movement trajectory that is not in the dataset, some unsupervised methods based on the action clustering algorithm will be used online. We hope that our work can improve the safety and reliability of HRC in the future.

REFERENCES

[1] G. Maeda, M. Ewerton, G. Neumann, R. Lioutikov, and J. Peters, "Phase estimation for fast action recognition and trajectory generation in human–robot collaboration," *Int. J. Robot. Res.*, vol. 36, nos. 13–14, pp. 1579–1594, Dec. 2017.

[2] M. Kalakrishnan, S. Chitta, E. Theodorou, P. Pastor, and S. Schaal, "STOMP: Stochastic trajectory optimization for motion planning," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2011, pp. 4569–4574.

[3] Q. Li, Y. Mu, Y. You, Z. Zhang, and C. Feng, "A hierarchical motion planning for mobile manipulator," *IEEJ Trans. Electr. Electron. Eng.*, vol. 15, no. 9, pp. 1390–1399, Sep. 2020, doi: 10.1002/tee.23206.

[4] O. S. Oguz, O. C. Sari, K. H. Dinh, and D. Wollherr, "Progressive stochastic motion planning for human-robot interaction," in *Proc. 26th IEEE Int. Symp. Robot Hum. Interact. Commun. (RO-MAN)*, Aug. 2017, pp. 1194–1201.

[5] V. Dutta and T. Zielinska, "Prognosing human activity using actions forecast and structured database," *IEEE Access*, vol. 8, pp. 6098–6116, 2020.

[6] A. Elnagar, "Prediction of moving objects in dynamic environments using Kalman filters," in *Proc. IEEE Int. Symp. Comput. Intell. Robot. Autom.*, Dec. 2001, pp. 414–419.

[7] A. Mogelmose, M. M. Trivedi, and T. B. Moeslund, "Trajectory analysis and prediction for improved pedestrian safety: Integrated framework and evaluations," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2015, pp. 330–335.

[8] Y. Maeda, T. Hara, and T. Arai, "Human-robot cooperative manipulation with motion estimation," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, vol. 4, Oct./Nov. 2001, pp. 2240–2245.

[9] B. Corteville, E. Aertbelien, H. Bruyninckx, J. De Schutter, and H. Van Brussel, "Human-inspired robot assistant for fast point-to-point movements," in *Proc. IEEE Int. Conf. Robot. Autom.*, Apr. 2007, pp. 3639–3644.

[10] B. Choo, M. Landau, M. DeVore, and P. Beling, "Statistical analysis-based error models for the microsoft KinectTM depth sensor," *Sensors*, vol. 14, no. 9, pp. 17430–17450, Sep. 2014.

[11] J. S. Park, C. Park, and D. Manocha, "I-planner: Intention-aware motion planning using learning-based human motion prediction," *Int. J. Robot. Res.*, vol. 38, no. 1, pp. 23–39, Jan. 2019.

[12] M. Wu, B. Taetz, E. D. Saraiva, G. Bleser, and S. Liu, "On-line motion prediction and adaptive control in human-robot handover tasks," in *Proc. IEEE Int. Conf. Adv. Robot. Social Impacts (ARSO)*, Oct. 2019, pp. 1–6.

[13] J. Mainprice and D. Berenson, "Human-robot collaborative manipulation planning using early prediction of human motion," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Nov. 2013, pp. 299–306.

[14] C. Perez-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. (ICRA)*, May 2015, pp. 6175–6182.

[15] S. Calinon, "A tutorial on task-parameterized movement learning and retrieval," *Intell. Service Robot.*, vol. 9, no. 1, pp. 1–29, Jan. 2016.

[16] R. Luo and D. Berenson, "A framework for unsupervised online human reaching motion recognition and early prediction," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2015, pp. 2426–2433.

[17] R. Quintero Minguez, I. Parra Alonso, D. Fernandez-Llorca, and M. A. Sotelo, "Pedestrian path, pose, and intention prediction through Gaussian process dynamical models and pedestrian activity recognition," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 5, pp. 1803–1814, May 2019.

[18] T. Fuse and K. Kamiya, "Statistical anomaly detection in human dynamics monitoring using a hierarchical Dirichlet process hidden Markov model," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 11, pp. 3083–3092, Nov. 2017.

[19] D. Vasquez, T. Fraichard, O. Aycard, and C. Laugier, "Intentional motion on-line learning and prediction," *Mach. Vis. Appl.*, vol. 19, nos. 5–6, pp. 411–425, Oct. 2008.

[20] 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.

[21] Z. Wang, P. Jensfelt, and J. Folkesson, "Modeling spatial-temporal dynamics of human movements for predicting future trajectories," in *Proc. AAAI Conf. Artif. Intell. Workshops*, 2015, pp. 42–48.

[22] Y. S. Razin, K. Pluckter, J. Ueda, and K. Feigh, "Predicting task intent from surface electromyography using layered hidden Markov models," *IEEE Robot. Autom. Lett.*, vol. 2, no. 2, pp. 1180–1185, Apr. 2017.

[23] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series," in *Proc. Int. Conf. Knowl. Discovery Data Mining*, 1994, pp. 359–370.

[24] F. Zhou and F. De la Torre, "Generalized time warping for multi-modal alignment of human motion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 1282–1289.

[25] P. A. Lasota and J. A. Shah, "Bayesian estimator for partial trajectory alignment," in *Proc. 15th Robot., Sci. Syst.*, vol. 15, Jun. 2019, pp. 1–9, doi: 10.15607/RSS.2019.XV.080.

[26] N. F. Duarte, M. Raković, J. Tasevski, M. I. Coco, A. Billard, and J. Santos-Victor, "Action anticipation: Reading the intentions of humans and robots," *IEEE Robot. Autom. Lett.*, vol. 3, no. 4, pp. 4132–4139, Oct. 2018.

[27] S. Trick, D. Koert, J. Peters, and C. A. Rothkopf, "Multimodal uncertainty reduction for intention recognition in human-robot interaction," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Nov. 2019, pp. 7009–7016.

[28] Q. Wang, W. Jiao, R. Yu, M. T. Johnson, and Y. Zhang, "Virtual reality robot-assisted welding based on human intention recognition," *IEEE Trans. Autom. Sci. Eng.*, vol. 17, no. 2, pp. 799–808, Apr. 2020.

[29] J. Lanini, H. Razavi, J. Urain, and A. Ijspeert, "Human intention detection as a multiclass classification problem: Application in physical human–robot interaction while walking," *IEEE Robot. Autom. Lett.*, vol. 3, no. 4, pp. 4171–4178, Oct. 2018.

[30] B. Xu, J. Li, Y. Wong, Q. Zhao, and M. S. Kankanhalli, "Interact as you intend: Intention-driven human-object interaction detection," *IEEE Trans. Multimedia*, vol. 22, no. 6, pp. 1423–1432, Jun. 2020.

[31] G. Ferrer and A. Sanfeliu, "Behavior estimation for a complete framework for human motion prediction in crowded environments," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Jun. 2014, pp. 5940–5945.

[32] R. Quintero, J. Almeida, D. F. Llorca, and M. A. Sotelo, "Pedestrian path prediction using body language traits," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2014, pp. 317–323.

[33] E. Rehder, F. Wirth, M. Lauer, and C. Stiller, "Pedestrian prediction by planning using deep neural networks," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 5903–5908.

[34] H. S. Koppula and A. Saxena, "Anticipating human activities using object affordances for reactive robotic response," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 1, pp. 14–29, Jan. 2016.
[35] B. Berret, E. Chiovetto, F. Nori, and T. Pozzo, "Evidence for composite cost functions in arm movement planning: An inverse optimal control approach," *PLOS Comput. Biol.*, vol. 7, no. 10, pp. 1–18, 2011, doi: 10.1371/journal.pcbi.1002183.
[36] J. Mainprice, R. Hayne, and D. Berenson, "Goal set inverse optimal control and iterative replanning for predicting human reaching motions in shared workspaces," *IEEE Trans. Robot.*, vol. 32, no. 4, pp. 897–908, Aug. 2016.
[37] O. S. Oguz, Z. Zhou, and D. Wollherr, "A hybrid framework for understanding and predicting human reaching motions," *Frontiers Robot. AI*, vol. 5, pp. 1–21, Mar. 2018, doi: 10.3389/frobt.2018.00027.
[38] H. Ben Amor, G. Neumann, S. Kamthe, O. Kroemer, and J. Peters, "Interaction primitives for human-robot cooperation tasks," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2014, pp. 2831–2837.
[39] G. J. Maeda, G. Neumann, M. Ewerton, R. Lioutikov, O. Kroemer, and J. Peters, "Probabilistic movement primitives for coordination of multiple human–robot collaborative tasks," *Auto. Robots*, vol. 41, no. 3, pp. 593–612, Mar. 2017.
[40] O. S. Oguz, V. Gabler, G. Huber, Z. Zhou, and D. Wollherr, "Hybrid human motion prediction for action selection within human-robot collaboration," in *Proc. Int. Symp. Exp. Robot.*, 2016, pp. 289–298.
[41] O. S. Oguz, B. M. Pfirrmann, M. Guo, and D. Wollherr, "Learning hand movement interaction control using RNNs: From HHI to HRI," *IEEE Robot. Autom. Lett.*, vol. 3, no. 4, pp. 4100–4107, Oct. 2018.
[42] P. Kratzer, M. Toussaint, and J. Mainprice, "Prediction of human full-body movements with motion optimization and recurrent neural networks," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Aug. 2020, pp. 1792–1798.
[43] G. Glonek and A. Wojciechowski, "Hybrid orientation based human limbs motion tracking method," *Sensors*, vol. 17, no. 12, p. 2857, 2017, doi: 10.3390/s17122857.
[44] D. McIlwraith, J. Pansiot, and G.-Z. Yang, "Wearable and ambient sensor fusion for the characterisation of human motion," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2010, pp. 5505–5510.
[45] N. Bu, M. Okamoto, and T. Tsuji, "A hybrid motion classification approach for EMG-based human–robot interfaces using Bayesian and neural networks," *IEEE Trans. Robot.*, vol. 25, no. 3, pp. 502–511, Jun. 2009.
[46] H. C. Ravichandar, A. Kumar, and A. P. Dani, "Gaze and motion information fusion for human intention inference," *Int. J. Intell. Robot. Appl.*, vol. 2, no. 2, pp. 136–148, 2018.
[47] I. Pasciuto, S. Ausejo, J. T. Celigüeta, Á. Suescun, and A. Cazón, "A hybrid dynamic motion prediction method for multibody digital human models based on a motion database and motion knowledge," *Multibody Syst. Dyn.*, vol. 32, no. 1, pp. 27–53, 2014.
[48] P. A. Lasota and J. A. Shah, "A multiple-predictor approach to human motion prediction," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Jun. 2017, pp. 2300–2307.
[49] Y. Xiang, J. S. Arora, and K. Abdel-Malek, "Hybrid predictive dynamics: A new approach to simulate human motion," *Multibody Syst. Dyn.*, vol. 28, no. 3, pp. 199–224, 2012.
[50] J. Cui, Y. Liu, Y. Xu, H. Zhao, and H. Zha, "Tracking generic human motion via fusion of low- and high-dimensional approaches," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 43, no. 4, pp. 996–1002, Jul. 2013.
[51] C. E. Rasmussen, "Gaussian processes in machine learning," in *Proc. Summer School Mach. Learn.* Berlin, Germany: Springer, 2003, pp. 63–71.
[52] H. Bijl, J.-W. van Wingerden, T. B. Schön, and M. Verhaegen, "Online sparse Gaussian process regression using FITC and PITC approximations," *IFAC-PapersOnLine*, vol. 48, no. 28, pp. 703–708, 2015.
[53] H. Bijl, T. B. Schon, J.-W. Wingerden, and M. Verhaegen, "System identification through online sparse Gaussian process regression with input noise," *IFAC J. Syst. Control*, vol. 2, pp. 1–11, Dec. 2016.
[54] A. Mchutchon and C. E. Rasmussen, "Gaussian process training with input noise," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 1341–1349.
[55] V. Dutta and T. Zielinska, "Predicting human actions taking into account object affordances," *J. Intell. Robot. Syst.*, vol. 93, no. 3, pp. 745–761, 2019.
[56] U. Pattacini, F. Nori, L. Natale, G. Metta, and G. Sandini, "An experimental evaluation of a novel minimum-jerk Cartesian controller for humanoid robots," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2010, pp. 1668–1674.

QINGHUA LI received the Ph.D. degree in signal and information processing from Shandong University, in 2009. He is currently an Associate Professor with the Qilu University of Technology (Shandong Academy of Sciences), Jinan, China. His current research interests include machine learning and machine vision.

ZHAO ZHANG (Graduate Student Member, IEEE) received the B.S. degree in communication engineering from the Qilu University of Technology (Shandong Academy of Sciences), Jinan, China, in 2017, where he is currently pursuing the M.S. degree with the School of Electrical Engineering and Automation. His research interests include machine learning, machine vision, and human robot collaboration.

YUE YOU (Graduate Student Member, IEEE) received the B.S. degree in electronic information engineering from the Qilu University of Technology, Jinan, China, in 2018, where she is currently pursuing the M.S. degree in control engineering. Her research interests include motion planning and particularly the intelligent path planning.

YAQI MU (Graduate Student Member, IEEE) received the B.S. degree in electrical engineering and automation from Huaiyin Normal University, Huaian, China, in 2018. She is currently pursuing the M.S. degree in control engineering with the Qilu University of Technology, Jinan, China. Her research interests include path planning, motion planning, and particularly the motion planning of mobile manipulators.

CHAO FENG received the Ph.D. degree in signal and information processing from Shandong University, in 2015. He is currently a Lecturer with the Qilu University of Technology (Shandong Academy of Sciences), Jinan, China. His current research interests include homomorphic encryption, remote data checking, and motion planning.