Real-Time Walking Gait Estimation for Construction Workers using a Single Wearable Inertial Measurement Unit (IMU)

Siyu Chen, Srikanth Sagar Bangaru, Tarik Yigit, Mitja Trkov, Chao Wang, and Jingang Yi

Abstract—Real-time gait detection and pose estimation are critical for safety monitoring and prevention of work-related musculoskeletal disorders for construction workers. We present a single wearable inertial measurement unit (IMU)-based gait detection and pose estimation for human walking on flat and sloped surfaces. The gait detection algorithm is built on a recurrent neural network-based method and its outcome is then used in the full-body pose estimation. The detection scheme also predicts the terrain slope information in real-time. The pose estimation is obtained through learned motion manifold in latent space with the Gaussian process dynamic model. Extensive experiments of different walking patterns and speeds on the level and sloped surfaces are conducted to validate and demonstrate the design. The proposed algorithm can detect gait activities with 96% accuracy, the estimated human pose errors are within 8.30 degs, and the detection latency is within 18.6 ms using only a single IMU attached to a human shank.

I. INTRODUCTION

Construction workers often suffer intense physical effort, and are exposed to serious safety and health risks in hazardous, dynamic working environments. One of the most common locomotions in construction trades is walking gait on flat and sloped surfaces (e.g., scaffold workers and roofers). It is critical to monitor workers’ gaits and body poses in real-time for safety and health conditions [1]. Wearable inertial measurement units (IMUs) are particularly attractive for gait detection and posture estimation in construction because they are small-size, low-cost and non-intrusive [2]. In [3], two IMUs were attached to the back of the helmet and the worker’s back for head, neck and trunk inclination estimation. The wearable sensors in [4] included eight IMUs on the trunk and limbs to detect gaits and motion of construction workers. In [5], 17 IMUs were used to identify poses of masonry workers using support vector machines. Comparison results of various IMU locations on the human body were reported in [6]. The pose estimation accuracy of the IMU measurements in these studies has however not been systematically and extensively validated. Moreover, wearable IMUs were not studied for real-time applications in the above-mentioned work and accurate limb poses were not among the focus.

In [7], a real-time gait detection was presented to capture walking gait events over level and inclined surfaces and staircases using a single IMU. Similar types of real-time walking gaits detection were also reported in [8]–[10] for periodic movement using machine learning methods. For gait detection of non-periodic human movements, [4] used a set of wearable IMUs on human limbs and trunk to monitor construction workers’ activities. Wearable IMU-based human gait detection was also presented in [11] for real-time applications.

As mentioned above, increasing use of machine learning techniques was reported in the recent years for human activity or gait classifications, and readers can refer to two reviews [12], [13]. In [14], recurrent neural network (RNN) with long short-term memory (LSTM) cells were used to reconstruct human poses during various motions in real time. A hierarchical multivariable hidden Markov model was employed in [15] for full-body locomotion reconstruction of human walking, running, jumping, and hopping motion on a flat floor using one IMU attached to the subject’s shank. Considering the periodic walking gait, a phase-functioned neural network model was presented in [16] to generate the human walking character in animation with a fast computational capability.

Inspired by these work and motivated by construction applications, we present a real-time gait detection and pose estimation scheme for walking on flat and sloped surfaces using one single wearable IMU. An LSTM approach is first used to detect the human walker gaits and the slope angle. A learned motion manifold is then constructed using the gait activity information. The pose estimation is built on the learned motion manifold and the IMU measurements. We use the Gaussian process dynamic model (GPDM) to construct the human motion manifold [17]. Due to periodic feature in biped walking, the learned GPDM is a closed manifold in latent space and similar to [16]. A phase variable is used to parameterize the GPDM model to predict the joint angles in real time. Extensive human experiments are conducted on flat and sloped surfaces that represent roof workers in construction. The experimental results demonstrate the efficacy and effectiveness of the design. The main contribution of the work lies in the novel integrated LSTM and GPDM gait detection and pose estimation for real-time applications using one single IMU. The design provides a potential enabling tool for wearable robotic systems to reduce the risk of work-related musculoskeletal disorders [18], [19].
II. EXPERIMENTS AND DATA COLLECTION

A. Experimental Setup

We created a laboratory environment that mimicked construction workers on the level and sloped surfaces. A wooden slope structure with glued anti-skid tapes was designed as a roof surface and the roof slope can be adjusted to reach any angle up to 40 degs. When walk cycle ends and foot strike is detected, the gait phase variable \( s \) is reset to the beginning of the stride, that is, \( s = 0 \).

Fig. 1 shows the overview of the data collection systems used in the experiments.

![Fig. 1. Laboratory experimental setup mimicking roof workers in construction. (a) Wearable IMU and embedded systems. (b) Walking setup on variable sloped wooden structure.](image)

To measure the human kinematics and motion, an IMU sensor (from WitMotion Inc.) was attached to the right shank of the subject; see Fig. 1(a). The raw IMU signal consists of 12 measurements (i.e., 3-joint angles, 3-gyroscopes, 3-linear acceleration, and 3-magnetometers). The IMU sensor was carefully placed on the same location of the shank segment with a fixed orientation before each trial. A motion capture system (8 Vantage cameras, Vicon Motion Systems Ltd.) was used to collect marker position for reference human motion at a sampling frequency of 100 Hz. A total of 39 markers were placed on subjects’ lower and upper limbs, trunk and head to represent full-body motion. Joint angles were calculated using custom algorithms in MATLAB software (Version R2020a, MathWorks Inc.) using the marker positions. The IMU measurements were sampled at 100 Hz, and the data were wirelessly transmitted to the portable embedded computer (Intel NUC7i7DNK, Intel Corp.) which was worn by the subject; see Fig. 1(a). In the training phase, all the data were connected and synchronized through a desktop computer and portable high-performance embedded computer. In the testing phase, data collection, real-time gait detection, and pose estimation algorithm were implemented on a desktop computer (Dell XPS-8953, Dell Inc. The desktop computer is equipped with Intel i7-8700 CPU, Nvidia 1030 GPU, and 16 GB RAM.

B. Experimental Protocol

Five healthy subjects \((n = 5, \text{ age: 30 } \pm 3 \text{ years, weight: } 73.3 \pm 6.5 \text{ kg, height: 172.0 } \pm 6.7 \text{ cm})\) were recruited for experiments, and they are capable of walking on flat and sloped surfaces. Four types of walking patterns were defined in experiments: walking on the level ground, walking up on the slope, walking transverse on the slope and walking down the slope; see Fig. 1(b). The subjects were instructed to use their normal gaits and self-selected speeds to walk on the level and sloped surfaces. An informed consent form was signed by all the subjects, and the Institutional Review Board (IRB) at Rutgers University approved the testing protocols.

The subjects started walking on the level ground surface and then followed the sequence of up the slope, transverse the slope and down the slope in the clockwise direction and finally returned to the starting location. The subjects repeated the sequence in the first four minutes for the training data collection phase, and then in the last minute they only walked transverse the slope from one side to the other side to collect the data in the symmetric direction. Three slope angles were selected in the training experiment: 5, 10, and 15 degs. Each data collection trial was repeated several times to obtain enough training data set. For real-time pose estimation tests, IMUs data were streamed live to a local computer and treated as inputs of the trained model.

C. Data Preprocess

The IMU and motion capture data were preprocessed by customized scripts. The walking activities were labeled based on the four different patterns. In each activity, only the complete strides were used for data analysis. The stride information was extracted from motion data, and stride length and walking speeds were calculated. The beginning and ending of each stride were defined by using the location of heel markers. The current and next right foot heel strikes were set as \( s = 0 \% \) and 100% of the gait phase, respectively, where \( s \) is the gait phase variable. To enlarge the training data set, multiple strides of the same activity were concatenated. To derive the latent space variables, averaged strides parameters from all the subjects were calculated. The strides from the same activity were resampled to be in the same frame length.

The corresponding IMU data were also concatenated for the same activities. IMU data were reshaped for the training purpose and 60 consecutive frames were combined as one frame data. The IMU sequence data was normalized with the mean and standard derivation of the entire data set, which was then used as inputs of the model. In the testing phase, 59 past frame with one current frame data were combined to form the input to the pose estimation model. The same mean and standard deviation were used to standardize and normalize the real-time IMU data, that is, the IMU sequence was subtracted by mean and divided by standard derivation of the training data.

The pre-trained learning models and neural network parameters were stored in the local computer. The real-time IMU sequence data were streamed into the learned model, and the subject’s motion activities and pose were estimated. In a real-time application, we detected foot strike by inspecting the sudden drop of the gait phase variable and the variation of the linear acceleration data. When walk cycle ends and foot strike is detected, the gait phase variable \( s \) is reset to the beginning of the stride, that is, \( s = 0 \).
III. LEARNING-BASED GAIT DETECTION AND POSE ESTIMATION

A. Overview

Fig. 2 illustrates the overview of the human walk gait detection and pose estimation scheme. The approach contains three tasks: walking activity detection, gait phase detection and full-body pose estimation. An LSTM-based classification and regression model is used to identify the walking activities, and the estimated slope angle (denoted as $\theta$) is also obtained. The gait propagation and phase variable $s \in [0,1]$ are estimated using a second LSTM-based RNN. Finally, a full-body human pose estimation scheme is obtained using the GPDM model. In the following, we describe each of these modules in details.

B. RNN-based Activity Detection and Gait Phase Estimation

The three layers in Fig. 3 show the schematic of the LSTM-based walking activity detection, slope angle estimation and gait phase estimation. The LSTM is a recurrent neural network architecture to learn sequential information using memory cells that store and output information to capture the temporal relationships. As shown in the left bottom of Fig. 3, the information update is through various gates and the relationships among the input gate, forget gate and output gate [20]. For the training purpose, the raw IMU data are collected and labeled for four walking patterns on level surface, up the slope, transverse the slope, and down the slope with different slope angles (i.e., $\theta = 0, 5, 10, 15$ degs). The IMU data are scaled and reshaped to three-dimensional input (i.e., samples, timesteps, features) for the LSTM models.

For the activity classification model, the single LSTM hidden layer has 50 neurons and the output of the LSTM layer is passed through a Dropout (0.2) layer to randomly drop 20% units from the network to prevent overfitting of the model. The outputs from the dropout layer are passed through a fully connected layer with 50 neurons with ReLU. Finally, the fully connected hidden layer is connected to the Softmax activation function, which converts the class scores to probabilities such that the activity with the highest probability can be recognized. The slope detection model consists of a two-stacked LSTM-based networks. The first LSTM hidden layer has 64 neurons followed by a Dropout (0.2) layer and the second LSTM and dropout layers are used to generate the output. For the classification model, the categorical cross-entropy and mean squared error are used as loss functions. Adam optimizer is used for both the activity classification and the slope angle estimation models. The model loss and accuracy curves are assessed to determine the model fit. The gait phase variable $s$ represents the current percentage of the stride, and is assumed to be increasing within one stride. The gait phase estimation model contains a two-stacked LSTM-based networks with one Dropout (0.2) layer between them, while each LSTM hidden layer has 32 neurons. Adam optimizer is used for the gait phase estimation models and mean squared error is used as the loss functions.

C. GPDM Model and Full-Body Pose Estimation

The periodic walking gaits in the high-dimensional joint angle space are represented in low-dimensional latent space through a manifold learning technique (i.e., GPDM). Let us denote the full-body joint angle as $y \in \mathbb{R}^D$ and the latent state variable as $x \in \mathbb{R}^d$, where $d$ and $D$ ($d \ll D$) are the dimensions of the latent space and the joint angle space, respectively. For each type of the $N$ walking activities $a_i$, $i = 1, \ldots, N$, and slope angle $\theta$, the latent dynamics for human motion are formulated as

$$
\mathcal{M}_i(\theta) : \begin{cases}
\frac{dx_i}{dt} = f_i(x_i, \alpha_i, u_i) + w_{pi}, \\
y_i = g_i(x_i, \beta_i, u_i) + w_{oi}
\end{cases}$$

(1)

where $x_i = x_i(s)$, $y = y_i(s)$, and $u_i = u_i(s)$ are latent state, joint angles and IMU measurements at gait phase variable $s$, respectively. $\alpha_i$ and $\beta_i$ are GP parameters and obtained from learning process, $w_{pi}$ and $w_{oi}$ are zero mean model noises for the state dynamics and the output models, respectively. In the training phase, the IMU data set $U_i = \{u_i\}_0^M$ and joint angle set $Y_i = \{y_i\}_0^M$ have number of $M$ training data points that are obtained for walking on surface with slope angle $\theta$. We then estimate the mappings $f_i(\cdot)$ and $g_i(\cdot)$ in (1) by identifying parameters $\alpha_i$ and $\beta_i$ through minimizing the posterior probability

$$
\mathcal{L}_i = -\ln p(X_i, \alpha_i, \beta_i|Y_i, U_i, \hat{X}_i),
$$

(2)

where $\hat{X}_i = \{\hat{x}_i\}_0^M$ is used to initialize $X$ in the optimization process (i.e., label of
Fig. 4. Schematic of the learned manifolds \( \mathcal{M}_i(\theta_1) \) and \( \mathcal{M}_i(\theta_2) \) that are obtained at slope angles \( \theta_1 \) and \( \theta_2 \) for activity \( a_i \). Manifold \( \mathcal{M}_i(\theta^*) \) is an estimated from interpolation from \( \mathcal{M}_i(\theta_1) \) and \( \mathcal{M}_i(\theta_2) \).

\[
X), \text{ and probability } p(X_i, \alpha_i, \beta_i|Y_i, U_i, \hat{X}_i) \propto p(Y_i|X_i, \beta_i)p(X_i|U_i, \alpha_i)p(X_i)p(\alpha_i)p(\beta_i).
\]

It is known that nearby points in the joint angle space are likely located close together in the latent space from manifold learning results [21]. Therefore, we consider that the level curves that belong the same activity share the same topological shapes in the latent space with small variations. Fig. 4 illustrates the above-discussed learned motion manifold concept. Manifolds \( \mathcal{M}_i(\theta_1) \) and \( \mathcal{M}_i(\theta_2) \) are obtained respectively by training the GPDM models with data sets at slope angles \( \theta_1 \) and \( \theta_2 \) for activity \( a_i \). To estimate the manifold \( \mathcal{M}_i(\theta^*) \) at any given slope angle \( \theta^* = \gamma\theta_1 + (1 - \gamma)\theta_2 \), where \( 0 < \gamma < 1 \) is weight factor, we interpolate from \( \mathcal{M}_i(\theta_1) \) and \( \mathcal{M}_i(\theta_2) \) as shown in Fig. 4. Once \( \mathcal{M}_i(\theta^*) \) is obtained from activity \( a_i \) on surface with slope angle \( \theta^* \), with the gait phase variable \( s \), latent state \( x_i(s) \) is predicted by the state dynamics and thus joint angles \( y_i(s) \) by the output equation of (1).

IV. EXPERIMENTAL RESULTS

We only consider upper- and lower-limb joint angles in the sagittal plane, and 12 joint angles are captured in the study. In the GPDM models, the latent space dimension is taken \( d = 3 \) and the dimension of the joint angle space is \( D = 12 \). The total number of walking activity is \( N = 4 \) (i.e., walking on level surface, up the slope, transverse the slope, and down the slope) and the slope angle is considered within 15 degs.

Fig. 5 shows the confusion matrix for walking activity detection results and the overall detection accuracy is 96.73\%. The model can successfully detect each individual activity with more than 95\% accuracy. Fig. 6 further illustrates the performance of activity predictions on a sample of a test data for four activities. The blue spikes in the figure indicate false classifications by the LSTM-model. Only occasional false classifications occur between walking on the level surface and walking up or down the slope, due to the close similarity among these types of gaits.

To validate the feasibility of our method, gait propagation and human joint angles are estimated and compared with actual values calculated from motion capture system. Twelve human joint angles from lower- and upper-limbs were selected for validation in our application, namely, left and right side hip, knee, ankle, shoulder, elbow and wrist limb joints. Fig. 7(a) shows the gait phase prediction from the LSTM model for five complete gait cycles of a subject walking on the level surface. A representative length of a gait cycle is marked in Fig. 7(a), as the gait phase variable \( s \) increases from 0 to 100\%. Gait phase prediction was less accurate at the start of each stride, and it however convalesces as the gait propagates. Fig. 7(b) illustrates three selected joint angles (knee, hip, and wrist) from all twelve joints to demonstrate the performance of the pose estimation. The estimation results match closely with the joints that are close to the IMU location; see knee angle results in the top plot. On the other hand, the estimation errors for the wrist joint (the bottom plot) show a slight increase in discrepancies.

Table I lists the overall mean absolute errors of all joint angles for each activity with respect to different slope angles. The joint angle errors were calculated from all 12 joints during
Fig. 7. Gait phase and joint angle profile of a representative subject during level surface walking. (a) Predicted gait phases (blue solid lines) compared with the actual gait phases (red dash-dotted lines). (b) Predicted (blue solid lines) and actual (red dash-dotted lines) angles of the knee ($\phi_{\text{knee}}$), hip ($\phi_{\text{hip}}$) and wrist ($\phi_{\text{wrist}}$) joints.

Fig. 8. Comparison of the joint angle error profiles for variable slope angles ($\theta = 5, 10, 15$ degs) during different activities. Walking (a) on flat surface, (b) up the slope, (c) transverse the slope, and (d) down the slope. Knee ($\epsilon_{\text{knee}}$), hip ($\epsilon_{\text{hip}}$) and ($\epsilon_{\text{wrist}}$) wrist mean absolute joint angle error profiles are shown on the top, middle, and bottom subfigures, respectively. The thick curves are the mean error profiles, while the shaded areas show the one standard deviation around the mean values of all subject experiments.

The largest estimation error is for the subjects walking down the slope, which might result from the fact that the gait during walking down the slope is similar to that on the flat surface when the slope angle is small as seen in Fig. 5. Table I also includes results of the joint angle estimation error for the walking gaits on two sloped surfaces (8 and 12 degs) which were obtained by using the manifold interpolation approach. Note that the data from these two tests were not present in the original training data. The average joint angle estimation errors are slightly higher than those for the “trained” slope angles in experiments. In Table I, we also list the accuracy of the slope angle estimation results. The average slope angle estimation error is around 3.4 degs and standard derivation is around 2.2 degs. The estimation accuracy is not high and one possible reason is that the training data set sizes are not large enough.
TABLE I

<table>
<thead>
<tr>
<th>Activity</th>
<th>Learned model</th>
<th>Interpolated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Slope Angle</td>
<td></td>
</tr>
<tr>
<td>Up Slope</td>
<td>6.88°±2.59°</td>
<td>6.70°±2.39°</td>
</tr>
<tr>
<td>Transverse</td>
<td>5.35°±2.50°</td>
<td>8.45°±2.49°</td>
</tr>
<tr>
<td>Down Slope</td>
<td>8.95°±3.01°</td>
<td>6.30°±2.91°</td>
</tr>
<tr>
<td>Level</td>
<td>5.39°</td>
<td>5.45°</td>
</tr>
<tr>
<td>Slope Angle Estimation</td>
<td>4.02°±2.59°</td>
<td>4.14°±1.58°</td>
</tr>
</tbody>
</table>

Furthermore, Fig. 8 shows the mean joint angle errors of three selected limb joints with their variances over different slope angles (θ = 5°, 10°, and 15°) for all subjects. The joint angle error profiles of walking on flat surface, up the slope, transverse the slope and down the slope are shown in Fig. 8(a) to 8(d), respectively. The knee and hip joint angles have large mean errors with high variance in certain portion, which might be due to the detection latency and larger ranges of motion of these joints compared to the wrist joint when the subjects walk up on the slope. The accuracy of the joint angle reconstruction using several IMUs for various non-walking 3D human motions in [14] was approximately 16 degs. The preliminary results in this paper show that the proposed approach achieved a similar accuracy for human walking pose estimation using a single IMU.

We evaluated the latency of our framework by feeding the pre-recorded unknown data offline to the network in real-time. The latency of activity detection is about 2.2 ms, the slope angle detection latency is about 2.58 ms, and the latency of pose estimation is about 13.82 ms. Therefore, the overall latency of the algorithm is approximately 18.6 ms. The joint angle estimation errors might result from several plausible sources. The training data quality and size might restrict the learning model accuracy. We only considered using mean values of gait cycles for each activity for full-body pose estimation. This requires a specialized tuning of the GPDM hyperparameters. Instead of using the manifold interpolation for any estimated slope angles, we might be able to further use topological constraints to enhance the GPDM model accuracy.

V. CONCLUSION

This paper implemented a real-time walking gait and pose estimation framework for construction workers on flat and sloped surfaces. A combination of the LSTM-based network and the GPDM was mainly used in the estimation scheme. The walking activities were predicted, and human pose was predicted with average joint angle errors within 8.30 degs. One attractive property of the proposed scheme was the real-time capability with 18.6 ms latency using a single IMU attached to the distal portion of the right fibula. We are currently working on improving the real-time performance and also integrating the design with wearable assistive robotic devices for construction workers.