Biomedical Signal Processing and Control 62 (2020) 102117

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Combining deep learning and model-based method using Bayesian Inference for walking speed estimation

Qian Yuyang, Yang Kaiming *, Zhu Yu, Wang Wei, Wan Chenhui

Department of Mechanical Engineering, Tsinghua University, Beijing, 100084, China

ARTICLE INFO

Keywords:
Walking speed estimation
Deep learning
Bayesian Inference
Self-paced treadmill

ABSTRACT

In this paper, deep learning and model-based method were combined using Bayesian Inference to realize high accuracy and good generalization capability stride-by-stride walking speed estimation by low cost inertial measurement unit (IMU) sensor. Long Short-Term Memory (LSTM) network was applied to train the prediction model because of its ability to consider the temporal correlation of multi-dimensional kinematic parameters during one stride. To improve the performance with unseen subjects, a model-based method was introduced for its relatively good generalization capability. Fusion strategy based on Bayesian Inference was applied to take advantage of the two methods which considered the estimation derived from different methods as abstract sensors. Six healthy young adults performed treadmill walking with shank-mounted IMU and the range of the walking speed was 2.5 km/h to 5 km/h at an increment of 0.5 km/h. Leave-one-subject-out (LOSO) cross-validation was performed to analyze the generalization capability. For deep learning method, the root mean square error (RMSE) of the model trained by all available subjects was 0.026 m/s and the RMSE of LOSO cross-validation was 0.066 m/s which indicated a low generalization capability. After the fusion strategy was applied, RMSE of the model trained by all available subjects was 0.023 m/s which was slightly improved, while the RMSE of LOSO cross-validation was reduced to 0.036 m/s which indicated that accuracy and the generalization capability was greatly improved. In addition, this accurate estimation can be easily realized online which is essential for locomotion interactive systems (e.g. self-paced treadmill).

1. Introduction

The most common and natural way for our daily travel is walking. Spatial and temporal gait parameters during walking have been extensively studied for healthy and pathological populations. Among these parameters, walking speed has been recognized as an important quantity in fitness applications, healthcare and rehabilitation [1–4]. Recently, self-paced treadmill walking has become increasing popular in gait assessment, training and rehabilitation as an alternative to traditional fixed-speed treadmill walking [5–9]. This kind of walking mode allows the subject to continually control and intrinsically select the walking speed, presumably leading to a more natural gait. High accuracy walking speed estimation is essential for self-paced treadmill walking because it is realized by positional feedback and walking speed feed-forward control strategy [10]. Motion capture systems perform well for accurate instantaneous walking speed estimation, but they are restricted to their size and cost for these applications. With the advances of MEMS, inertial measurement units (IMU) have become more and more popular in physical activity analysis because of the portability, improved performance, and low cost.

Walking speed estimation methods based on IMU fall into three categories: direct integration method, model-based method and machine learning method [11]. Comparing to the model-based methods and learning methods, direct integration methods are free from subject-specific model calibration or training progress. One important assumption for this kind of methods was the zero-velocity update (ZUPT), dealing with the drift effect. In 2005, Sabatini et al. [12] used a foot-mounted IMU and the foot-flat (FF) event was detected by the ZUPT, which achieved an RMSE of 0.18 km/h on treadmill walking at the speed range of 3 km/h to 6 km/h. In 2014, Mannini et al. [13] applied hidden Markov model based method and the best results achieved 2% RMSE during over-ground walking. Although a new sophisticated algorithm based on drift estimation has been developed recently [14], the estimation accuracy still relied on sensor performance. In general, direct integration methods were susceptible to senor noise and bias because of the dual integration, which was unsuitable for low-cost sensors.

For model-based method, walking speed is estimated by predefined human gait model. The accuracy depends on the validity of

* Corresponding author.
E-mail address: yangkm@mail.tsinghua.edu.cn (Yang K.).

https://doi.org/10.1016/j.bspc.2020.102117
Received 30 December 2019; Received in revised form 13 July 2020; Accepted 1 August 2020
Available online 11 August 2020
1746-8094/

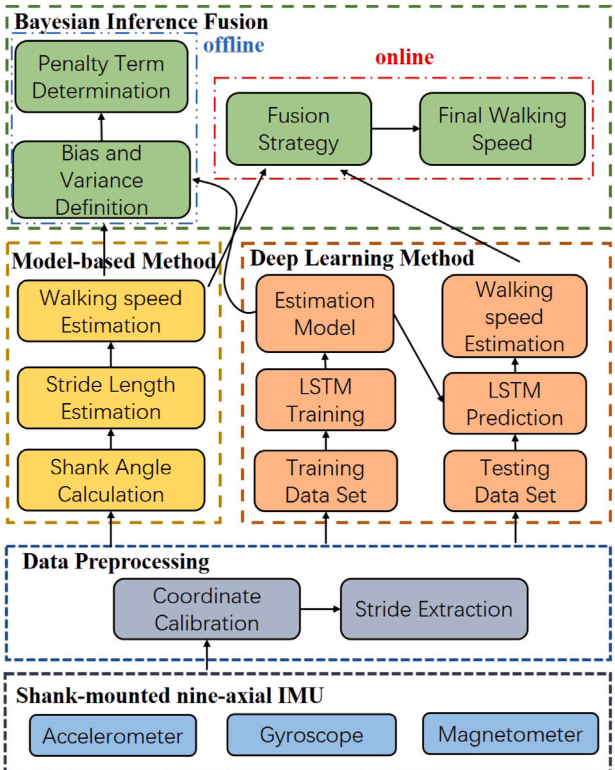

Fig. 1. System Architecture of the proposed fusion strategy based on Bayesian Inference.

the gait model and the gait model directly affects the complexity of the algorithm [15–17]. In 2011, Chen, S. et al. [15] proposed a refined pendulum gait model and the walking speed was estimated by a single shank-mounted IMU. In this model, the two legs were assumed symmetrical, each leg was modeled as two segments (thigh and shank) during the backward swing phase (the hip rotation angle was neglected in this phase) and one segment during forward swing phase. This method achieved an RMSE of 0.095 m/s during treadmill walking at speed ranging from 1 to 3 MPH. In 2002, Aminian et al. [16] solved the complete gait model with separate shank and thing segments with the same assumption of symmetry between two legs. The overall RMSE was 0.06 m/s (6.7%) for walking speed estimation. In 2013, Jwu-Sheng Hu et al. [17] proposed a virtual inverted pendulum which was derived from the rolling-foot model and considered the waist rotation effect. Experimental results showed a 0.58% absolute error and 0.72% error deviation. However, the kinematic model of this method was extremely complexity. Carefully human model parameters calibration was needed to achieve relatively high accuracy and the experiment was only performed on over-ground walking rather than varying walking speed by treadmill walking.

Machine learning methods have been widely studied recently years, plenty of algorithms have been introduced to improve the performance, such as support vector machine, artificial neural network and deep convolutional neural network [18–21]. In 2017, McGinnis et al. utilized wearable accelerometer arrays mounted on several parts of body, such as shank, sacrum, and thigh [19] and achieved an RMSE of 0.11 m/s. In 2018, Shrestha et al. [20] applied deep convolutional neural network to estimate the walking speed by smartphone sensors and the overall RMSE was 0.16 m/s. Thanks to the deep learning method, the result was quite comparable with the result obtained by several skin-mounted accelerometer sensors [19]. However, the machine learning methods would show low intra-subject variability and high inter-subject variability [21], and the generalization capability must be carefully considered.

In this paper, Long Short-Term Memory (LSTM) network was applied to automatically extract high-level temporal features of multi-dimensional kinematic parameters during one stride and fully connected layers were followed for regression. It was expected that the LSTM neural network model can achieve higher accuracy due to the advantages of deep learning. However, the bias–variance dilemma [22] is inevitable, especially for human gait parameters estimation. The training dataset is always finite but the variety of human gait is infinite, thus the datasets are not subject to the independent identical distribution. As the training progress going on, the bias would decrease while the variance would increase rapidly. As a consequence, the output would be sensitive to the small changes of input. When the data from unseen subjects are fed to the network, the error would be uncontrollable. So, the high accuracy and good generalization capability for walking speed estimation are hard to achieve at the same time by learning methods. In contrast to the learning methods, model-based methods can achieve good generalization capability if the sensor was well calibrated and human parameters were well estimated.

In conclusion, with the advances of recently developed deep learning methods, high accuracy estimation can be expected, but the generalization capability should be carefully considered. Methods that combining deep learning methods and simple model-based methods may achieve both high accuracy and good generalization capability with single low-cost IMU senor. To take advantage of the learning methods and model-based methods, a novel fusion strategy based on Bayesian Inference was proposed in this paper. Firstly, LSTM network was applied to train the dataset to get high accuracy estimation because of its ability to consider the long-term temporal corrections of multi-dimensional kinematic parameters during one stride. Drop layer was followed to reduce over-fitting and fully connect layers were used for final regression. Then, the simple refined pendulum gait model [12] was introduced to estimate the walking speed individually. Finally, the results of the two methods derived from the same raw signals were considered as two abstract sensors. Biases and variances of the two abstract sensors were predefined by the results of training dataset and Bayesian Inference was introduced to fuse the two abstract sensors in order to get both high accuracy and good generalization capability. Penalty term was also introduced to improve the performance with the situations when deep learning method performed poorly for unseen subjects.

2. Method

2.1. System architecture

Fig. 1 illustrated the overall structure of the system. A shank-mounted IMU was used to collect the kinematic parameters of each subject during treadmill walking at different speeds. Firstly, the inertial measures were transferred from the sensor coordinate to the treadmill coordinate by quaternion rotation. Then, threshold-based local peak detection was applied for the angle velocity perpendicular to the sagittal plane to detect the FF event and extract stride cycles. The dataset was divided into two segments for training and prediction and the LSTM neural network was utilized to train the prediction model. Then, a simple refined pendulum human gait model was applied to estimate the walking speed individually. According the fusion strategy, the biases and variances of the two methods was predefined by the training dataset, as well as the parameters of the penalty term. Finally, the walking speed estimated by LSTM network and model-based method were combined by the fusion strategy for both high accuracy and good generalization capability estimation.

2.2. Data preprocessing

To calibrate the sensor mounting errors, the signals must be correctly transferred into the treadmill coordinate($\boldsymbol{O}$), axes xyz: x pointed forward, opposite to the moving direction of the treadmill, z pointed upward vertically and y perpendicular to x and z. There are anther two coordinates: the senor coordinate($\boldsymbol{B}$) and the global coordinate($\boldsymbol{G}$). The mounting calibration was used to transfer the measures in the sensor coordinate to the treadmill coordinate by the medium of the global coordinate. Firstly, the sensor was placed aligning with the treadmill coordinate for 30 s and the mean value of quaternion $\boldsymbol{q}_0$ was recorded, which represented the transformation relation from global coordinate to the treadmill coordinate.

$$q_0 = [q_0, q_1, q_2, q_3] \tag{1}$$

The measurements of IMU (i.e., accelerometer, gyroscope) at ith discrete time in the sensor coordinate can be expressed as follows:

$$\boldsymbol{a}_i^B = [a_x, a_y, a_z]^T \tag{2}$$

$$\boldsymbol{g}_i^B = [g_x, g_y, g_z]^T \tag{3}$$

Where $\mathbf{a}_i^B$, $\mathbf{g}_i^B$ represented the accelerations vector and gyroscope vector in the sensor coordinate at the discrete time i, a_x, a_y, a_z indicated the measures of tri-axial accelerometer, while and g_x, g_y, g_z indicated the measures of tri-axial gyroscope. The vectors in sensor coordinate can be easily transferred to global coordinate by quaternion rotation.

$$\boldsymbol{a}_i^G = \boldsymbol{q}_i^{-1} \circ \boldsymbol{a}_i^B \circ \boldsymbol{q}_i \tag{4}$$

$$\boldsymbol{g}_i^G = \boldsymbol{q}_i^{-1} \circ \boldsymbol{g}_i^B \circ \boldsymbol{q}_i \tag{5}$$

The measurement vectors in treadmill coordinate can be expressed as follows:

$$\boldsymbol{a}_i^O = \boldsymbol{q_0} \circ \boldsymbol{a}_i^G \circ \boldsymbol{q}_0^{-1} \tag{6}$$

$$\boldsymbol{g}_i^O = \boldsymbol{q_0} \circ \boldsymbol{g}_i^G \circ \boldsymbol{q}_0^{-1} \tag{7}$$

After the mounting calibration, gait cycles needed to be extracted for the purpose to obtain spatial and temporal gait parameters such as gait phase, step time, step length and walking speed. Based on the assumption that the shank angular velocity in the sagittal plane was close to zero during the FF event, local peak detection algorithm was applied to detect the FF event. To suppress the ripples of signals, a zero-phase, 3rd order Butterworth low-pass filter with a cutoff frequency of 3 Hz was applied. The cut-off frequency is determined by the fact that the main frequency components of gyroscope signals lie below 3 Hz. The time points of the FF events were recorded and the gait cycles were extracted based on the recorded time points.

2.3. Deep learning method

LSTM neural network can extract both long-term and short-term features from the raw data, presumably leading to better prediction accuracy. The training data D fed to the network can be expressed as follows:

$$\boldsymbol{Acc}_i = [\boldsymbol{a}_1^O, \boldsymbol{a}_2^O, \ldots, \boldsymbol{a}_M^O]^T \tag{8}$$

$$\boldsymbol{Gyr}_i = [\boldsymbol{g}_1^O, \boldsymbol{g}_2^O, \ldots, \boldsymbol{g}_M^O]^T \tag{9}$$

$$D = ([\boldsymbol{Acc}_i,\ \boldsymbol{Gyr}_i],\ y_i),\ i = 1, \ldots, N \tag{10}$$

Where the segment size M was fixed to 100 in this study which indicated that the number points was normalized to 100 for each extracted gait cycle. y_i represented the ground-truth walking speed of the ith stride and D was the training input for LSTM neural network walking speed estimation model.

LSTM model consists of a series of LSTM cells. As was shown in Fig. 2, LSTM cell at time t includes three gates (forget, input and output) and three vectors: $\mathbf{x}_t$ (input vector), $\mathbf{h}_t$ (recurrent hidden state) and $\mathbf{C}_t$ (long term state). A matrix of 6×M consisted of tri-axial accelerometer and tri-axial gyroscope measures of one stride was taken as the input sequence and fed to the cell sequentially. The outputs of the three gates at time t were calculated as Eq. (11) - Eq. (13). For t=0, both the previous cell state and hidden layer were set to zero vectors:

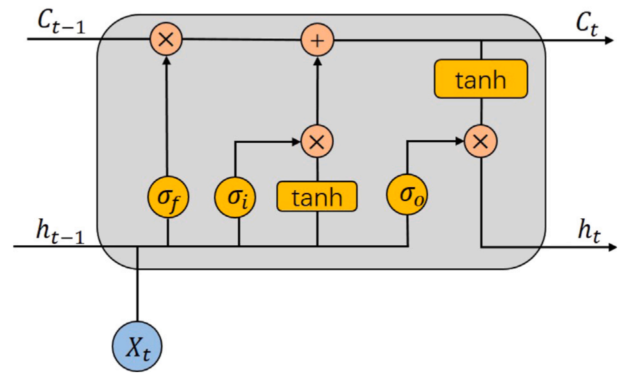

Fig. 2. Example for a LSTM cell.

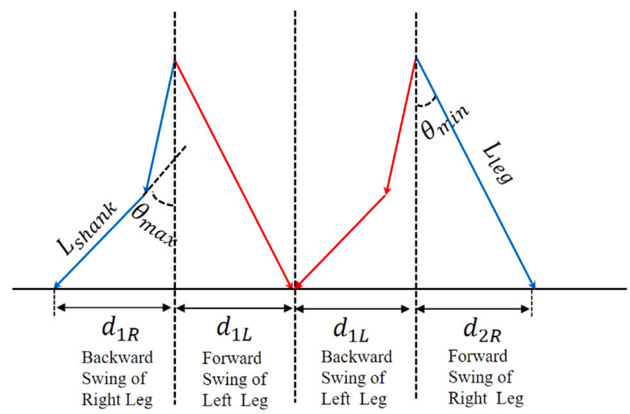

Fig. 3. Refined pendulum gait model for walking speed estimation.

$$\boldsymbol{f}_t = \sigma(\boldsymbol{W}_f \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t + \boldsymbol{b}_f]) \tag{11}$$

$$\boldsymbol{i}_t = \sigma(\boldsymbol{W}_i \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t + \boldsymbol{b}_i]) \tag{12}$$

$$\boldsymbol{o}_t = \sigma(\boldsymbol{W}_o \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t + \boldsymbol{b}_o]) \tag{13}$$

where W was the weight matrix, b was the bias matrix and $\sigma(\cdot)$ was a sigmoid function providing nonlinearity for gates. Then the candidate state value $\widetilde{\boldsymbol{C}}_t$ was calculated as Eq. (14), and the long-term state $\boldsymbol{C}_t$ was updated as follows:

$$\widetilde{\boldsymbol{C}}_t = tanh(\boldsymbol{W}_c \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_c) \tag{14}$$

$$\boldsymbol{C}_t = \boldsymbol{f}_t \circ \boldsymbol{C}_{t-1} + \boldsymbol{i}_t \circ \widetilde{\boldsymbol{C}}_t \tag{15}$$

The output of time-step t was calculated as Eq. (16), and only the last cell output was used for finally estimation.

$$\boldsymbol{h}_t = \boldsymbol{o}_t \circ tanh(\boldsymbol{C}_t) \tag{16}$$

The architecture and layer parameters of the LSTM neural network model were shown in Fig. 4. As was illustrated, the LSTM walking speed estimation model toke accelerometer and gyroscope measures as the input sequences, and the number of hidden layers was set to 100, the same as the number of the input sequence. The hidden dense layers had Rectified Linear Unit (ReLU) as their activation function, and the prediction dense layer has no activation function cause it is a regression task. The LSTM layers mapped the input sequences to an embedding in temporal feature space. Then, three cascading fully connected layers

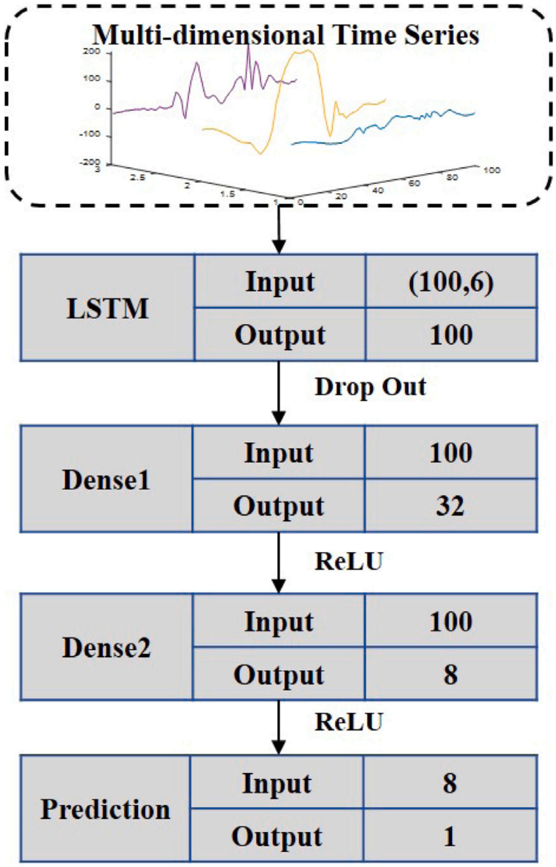

Fig. 4. Network structure for LSTM walking speed estimation model.

were added to build a map between feature space and target walking speed.

2.4. Model-based method

In this paper, the simple refined pendulum gait model [15] was chosen to be the gait model for walking speed estimation. Strides were divided by FF events as mentioned in the data preprocessing section, and stride time can be obtained as well. Step length can be computed by the gait model illustrated in Fig. 3. The distance from the center to fore leg and back leg during one step can be defined as:

$$d_{1R} = sin(\theta_{max}) \times L_{shank} \tag{17}$$

$$d_{2R} = sin(\theta_{min}) \times L_{Leg} \tag{18}$$

Where L_{shank} and L_{Leg} referred to the length measured from ankle joint to knee joint and the length from ankle joint to hip joint respectively. θ_{max} represented the maximum shank angle at initial swing and θ_{min} represented the minimum shank angle at the heel-strike events during forward swing. Assuming the two steps during one stride as symmetric, the stride length and walking speed can be calculated as:

$$SL = 2(d_{1R} + d_{2R}) \tag{19}$$

$$S_M = SL/ST \tag{20}$$

Where SL and ST referred to stride length and stride time respectively and S_M represented the walking speed estimated by the model-based method. The shank angles are calculated by the integration of angular velocity measured by gyroscope sensor signals in the sagittal plane:

$$\theta(n) = \theta(n-1) + \frac{T_P}{2}[\omega(n-1) + \omega(n)] \tag{21}$$

Where ω was the angular velocity obtained by gyroscope signals in sagittal plane, T_p was the sample time of the sensor (10 ms in this case) and θ(n) represented the dynamic shank angle at discrete time n. The swing range can be determined by differencing the maximum and minimum of the shank angle during one gait cycle.

It must be noted, although the refined pendulum gait model considers the leg as two segments, the angle of thigh is neglected because only one IMU is mounted and the thigh angle is unavailable in this study. As a consequence, errors exist definitely and must be considered, so as other human gait models. In this case, walking speed tends to be overestimated during low-speed region and underestimated during high-speed region. At very low speeds, the thigh tends to swing forward ahead of the plumb line so as to maintain a very short step length on the treadmill, resulting in a step length that is shorter than predicted, and vice versa at high speeds.

2.5. Fusion strategy based on Bayesian inference

In order to get high accuracy walking speed estimation as well as good generalization ability, deep learning method and model-based method were fused under the frame of Bayesian Inference. The two models are considered as two abstract sensors, the results of each sensor were assumed to satisfy Gaussian distribution. According to Bayesian Inference, the joint probability distribution can be calculated as follows:

$$P(x|x_1, x_2) = \frac{P(x_2|x, x_1)P(x|x_1)}{P(x_2|x_1)} \tag{22}$$

Where x_1, x_2 were measurements of the two sensors and x represented the true state of the observed parameter. $P(\cdot \mid \cdot)$ was the conditional probability and $P(x|x_1, x_2)$ represented the estimated distribution based on the two independent observations x_1, x_2. Given the two sensors were independent of each other, the probability distribution of the true value can be expressed as follows:

$$\begin{aligned} P(x|x_1, x_2) &= \frac{P(x_2|x)P(x|x_1)}{P(x_2)} \\ &= \frac{P(x|x_2)P(x|x_1)}{P(x)} \\ &\propto P(x|x_1)P(x|x_2) \end{aligned} \tag{23}$$

According to Eq. (23), the joint probability distribution of the true state was proportional to the product of the two sensors' probability distribution. According to the Gaussian distribution assumption, the joint probability distribution of true state can be calculated as follows:

$$\begin{aligned} P(x|x_1)P(x|x_2) &= \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(x-x_1)^2}{2\sigma_2^2}} \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{(x-x_2)^2}{2\sigma_2^2}} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\bar{x})^2}{2\sigma_1^2}} \end{aligned} \tag{24}$$

Applying the concept to the current issue, the biases and variances of the two abstract sensors needed to be determined. For deep learning method, the estimation was almost unbiased and the definition of variance was also easy to be done. The mean–variance of different ground truth walking speed was suitable for any walking speed as it would not differ a lot at different ground truth walking speed. However, for the model-based method applied in the current study, the mean value of the estimated walking speed was not around the ground truth and the bias error depends on the ground truth walking speed. As was mentioned in Section 2.5, the thigh angle was neglected during swing backward phase. At very low speeds, the thigh tends to swing forward ahead of plumb line so as to maintain a very short step length on the treadmill, resulting in a step length that is shorter than predicted, and vice versa at high speeds. As was shown in Fig. 5a, the deviation was near linear relationship with true walking speed. In order to correct the bias error, a linear fit was applied to determine the mean deviation with

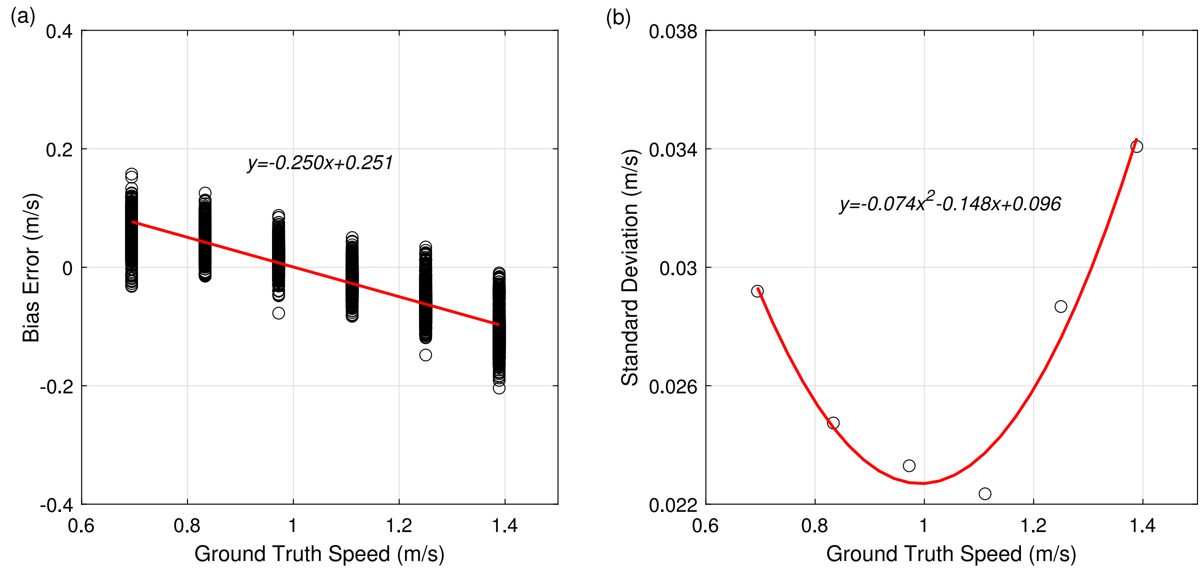

Fig. 5. (a) Bias error for model-based method. Bias error was scattered with respect to the ground truth speed and the red line was the linear fitting line of the bias error. (b) Variance for model-based method. The variance at each speed is scattered in the figure and the red line was quadratic fitting curve of the variance.

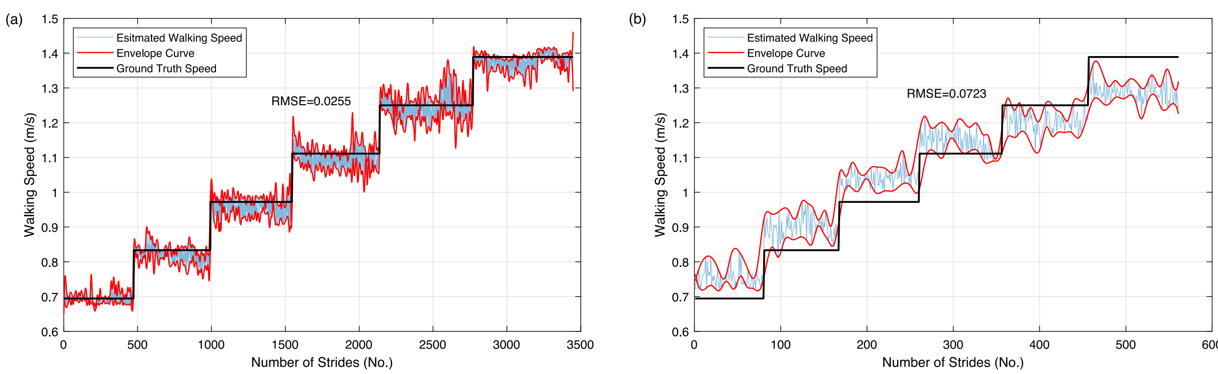

Fig. 6. (a) Estimated walking speed of the LSTM model trained by all available subjects. (b) Estimated walking speed of the LSTM model trained by LOSO cross-validation(take subject No.3 as an example).

respect to the ground truth walking speed. As was shown in Fig. 5b, the variance distribution was similar to U-shaped and the minimum variance is around 1 m/s, when the hip angles at toe-off were close to zero. Thus, the quadratic fit was applied to determine the variance of model-based method. It should be noted that the determination of bias and variance depends on the characteristic of the selected model, the processing above were only suitable for the model applied in the current study. So far, the variances and deviations of the two abstract sensors have been determined, the fused speed can be expressed as follows:

$$\alpha = \frac{\sigma_D^2}{\sigma_D^2 + \sigma_M^2(x_M)} \tag{25}$$

$$X = (1-\alpha)(x_M - E_M(x_M)) + \alpha x_D \tag{26}$$

Where X represented the fused walking speed, x_M, x_D were the estimated walking speed derived from model-based method and deep learning method individually, σ_M^2 and E_M are the predefined variance and deviation function of model-based method and σ_D^2 represents the mean–variance of deep learning method. It should be noted that the variances and deviations of the specific model were pre-determined as the function as true speed, but in Eqs. (25) - (26) they were determined as the function of estimation results derived from model-based method. This little error can be ignored because model-based has good generalization if the IMU sensor was well calibrated, the estimation results of the specific model are close to the true speed even the subject was not included in the dataset used for the pre-determination of the biases and variances.

For deep learning method, the variances and deviations could be totally different for subjects out of the training data. For the purpose to reduce the influence, the penalty term is introduced. If the estimation results of deep learning were far from the estimation of specific model (bias corrected), the reliability of deep learning method should be reduced according to the relative deviation percent between the two measurements. Therefore, the fusion strategy should be modified as follows:

$$\xi = \frac{|x_D - (x_M - E(x_M))|}{2(x_M - E(x_M))} \tag{27}$$

$$\alpha' = \alpha * max(0, (1-\xi)) \tag{28}$$

$$X = (1-\alpha')X_M + \alpha' x_D \tag{29}$$

Where ξ represented the penalty term and α' is the weighting factor of deep learning method. If the relative deviation percent between deep learning result and bias-corrected model-based method exceeded 50%, the weighting factor of deep learning method would be set to zero, which means the result derived from deep learning method was almost unreliable. The situation above may occur when unseen subjects or unseen speed were fed to the network.

Table 1
The root mean square errors (m/s) for all the subjects using different methods.

Subject	Model-based	Deep learning$_{in}$	Deep learning$_{out}$	Fusion strategy$_{in}$	Fusion strategy$_{out}$
1	0.037	0.027	0.087	0.024	0.030
2	0.075	0.026	0.061	0.023	0.033
3	0.061	0.030	0.072	0.022	0.038
4	0.074	0.029	0.070	0.020	0.039
5	0.060	0.022	0.046	0.023	0.043
6	0.065	0.027	0.050	0.023	0.031
Overall	0.068	0.026	0.066	0.023	0.036

The subscript *in* represented the model trained by all available subjects and the subscript *out* represented the model trained by LOSO cross-validation.

3. Experimental evaluation

3.1. Data acquisition

Six healthy young subjects, self-reported to be free from known neuromuscular disorders, gave informed consent to participate in this study. All subjects had prior experience walking on a treadmill. The subjects ranged in age from 21 to 26 years (24 ± 1.84) with an average height of 1.71 ± 0.05 m and a mass of 67.83 ± 7.30 kg. All participants signed institutionally approved consent forms before participating. The protocol was approved by the University Research Ethics Board.

An IMU (BWT901CL, Wit-Motion, China) was mounted on the shank of the participants to obtain three-dimensional accelerations and angular velocity and the sensor quaternion during treadmill walking. The quaternion was calculated by the sensor based on Kalman Filter and transferred to the recording device. The participants were first required to stand still for 10 s on the treadmill to get the configuration data. Then, the treadmill speed was set to 2.5 km/h to 5 km/h at an increment of 0.5 km/h. At each speed, the participants toke a 2-minute treadmill walking, the first 30-second walking was used to make the participants familiar to the walking speed and only the data after the 30 s were used for further analysis. Between each walking trial, a 2-minute rest time was taken by each participant to avoid the learning effect. Finally, a shank-mounted IMU dataset for walking speed estimation consisting of six subjects with the speed of [2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0] km/h was established. The strides for each subject at each walking speed were randomly divided into two parts. The number of strides for each part was equally, one part was used for training progress and the other for validation.

3.2. Evaluation metrics

The values of estimated walking speed at each detected stride were represented by $v(m,n,l)$. The index l runs over the number of strides L which was the total number of strides walked by the $n-th$ subject at $m-th$ walking speed (n=1,2,..., N, m=1,2,..., M; where M = N=6 in the current study). There are several standard metrics for learning methods such as false positive rate and false negative rate [23]. However, these metrics are more suitable for classification problems rather than estimation. In the current study, the performance was assessed by the following metrics. The mean value of estimated walking speed for $n-th$ subject at $i-th$ speed was represented by $M(m,n)$. The mean estimated speed was compared with the walking speed reference values v_{ref}, yielding the estimation error $E(m,n)$ for each walking trial. The estimated walking speed $v(m,n,l)$ were compared to the mean value $M(m,n)$ yielding the variance $V(m,n)$ for each walking trial:

$$M(m,n)=\frac{1}{L}\sum_{l=1}^{L_{m,n}}v(m,n,l) \tag{30}$$

$$E(m,n)=M(m,n)-v_{ref}(m) \tag{31}$$

$$V(m,n)=\frac{1}{L}\sum_{l=1}^{L_{m,n}}[v(m,n,l)-M(m,n)]^2 \tag{32}$$

The root mean square error (RMSE) was calculated for each detected stride:

$$RMSE=\sqrt{\frac{1}{LMN}\sum_{n=1}^{N}\sum_{m=1}^{M}\sum_{l=1}^{L}(v(m,n,l)-v_{ref}(m))^2} \tag{33}$$

For the purpose to evaluate the generalization capability, the LSTM network model and the fusion strategy were validated using a leave-one-subject-out (LOSO) cross-validation approach [13]. This validation approach trains the model using data from all subjects but one and testing the model on the subject that has not been included in the training dataset. The procedure was repeated for all available subjects. After comparing with the results with the model trained by all available subjects, the generalization capability can be evaluated by the ratio between the two results:

$$R=\frac{RMSE_{out}}{RMSE_{in}} \tag{34}$$

Where RMSE$_{in}$ represents the RMSE of the model trained by all subjects and RMSE$_{out}$ represents the average RMSE of the model trained by the LOSO cross-validation. It should be noted that the model for all available subjects was trained by half of the strides of each subject at every walking speed, and the results were tested by the rest of the strides. This kind of validation can assess how well the model can generalize to unseen subjects and is suited to deal with the processing of human motion data.

In general, values of the accuracy performance metrics close to zero would indicate better performing methods while the generalization capability metric should be close to 1. The mean error value carries information about the bias error relative to the reference for $n-th$ subject at $m-th$ speed and the variance represents the dispersion degree of the estimated walking speed. Given the mean error is a signed error measure, underestimations or overestimations that might occur during the walking trial could hide each other in the averaging progress involved in its calculation. This issue can be corrected by taking RMSE into account. In addition, to assess the generalization capability, the R-ratio was introduced.

3.3. Results of experimental study

RMSE of the shank-mounted IMU walking speed estimation for each subject using different method has been shown in Table 1. For the refined pendulum model used in the current study, the average RMSE was 0.068 m/s which was close to the LSTM model using LOSO cross-validation. As was mentioned above, this method trends to underestimation in high-speed region and overestimation in low-speed region. The deviations of each stride were scattered in Fig. 5a which trend to be linear with the ground truth, and the variances at each speed were scattered in Fig. 5b which trend to be quadratic with the ground truth. Although the results were not competitive with the LSTM model trained by all available subjects, the generalization capability of model-based method was better than learning methods. The estimation accuracy only depends on the validity of the gait model and the lower limb parameters.

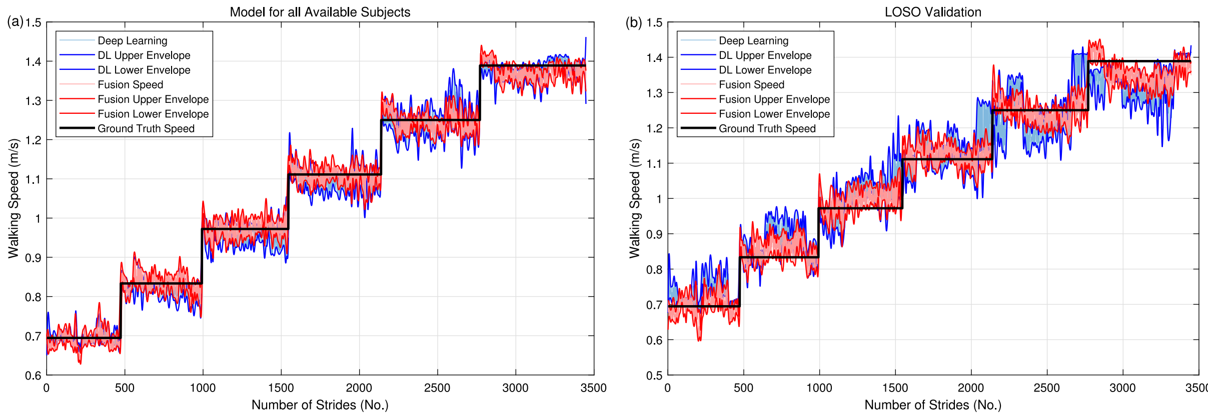

Fig. 7. (a) Comparison of the estimated walking speed trained by all available subjects derived from LSTM model and fusion strategy respectively. (b) Comparison of the estimated walking speed trained by LOSO cross-validation derived from LSTM model and fusion strategy respectively.

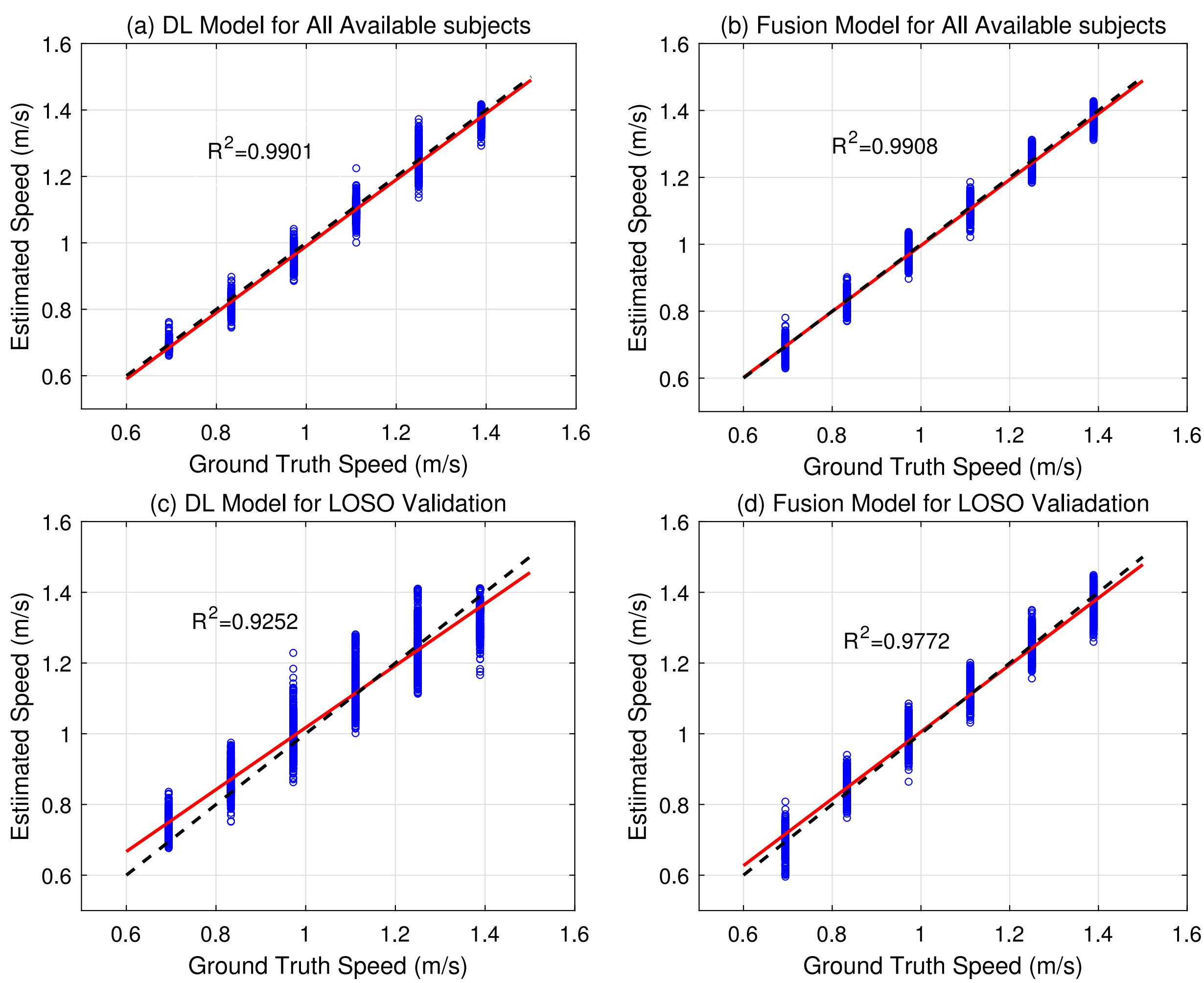

Fig. 8. Linear regression and the determination coefficient of different methods. The red line represented the best fit line and the dashed black line was a proportional function with the slope of 1.

For deep learning method, the overall RMSE of the LSTM model trained by all available subjects was 0.026 m/s, the estimated walking speed and its envelope curve were plotted in Fig. 6a. However, the average RMSE was increased significantly (0.066 m/s) when LOSO validation was applied to the LSTM model. Even so, this result was better than the result utilizing wearable accelerometer arrays mounted on several parts of body based on support vector machine (0.11 m/s) [19] and was comparable with the complex model-based method (0.06 m/s) [16]. The R-ratio of the LSTM method was 2.54 which indicated the generalization capability for deep learning method was very low. Take subject No.3 as an example, the RMSE of estimated walking speed using LOSO cross-validation was 0.072 m/s, and the estimated walking speed fluctuated a lot around the ground truth (Fig. 6b). Human gait characteristics vary from person to person, if the gait features have not been learned by the network, the accuracy of the estimated walking

Table 2
The parameters of best fit line for different methods.

Method	Best fit line
Deep learning$_{in}$	y=1.0002x − 0.0101
Fusion strategy$_{in}$	y=0.9855x + 0.0107
Deep learning$_{out}$	y=0.8765x + 0.1409
Fusion strategy$_{out}$	y=0.9461x + 0.0592

Table 3
Time efficiency of LSTM network.

Trainable parameters	Training epochs	Training time	Testing time
46324	400	48 min 50 s	3.884 s

speed would decrease a lot and the error could even be uncontrollable for some abnormal gait.

From the results above we can conclude that the deep learning method can get relatively high accuracy walking speed estimation while the generalization capability should be carefully considered. On the contrary, the model-based method could achieve better generalization capability. In order to take advantage of the two methods, fusion strategy based on Bayesian Inference was applied for each stride. The biases and variances of model-based method were determined by fitting approach (red curves in Fig. 5) and the variance of the LSTM neural network method was determined by the average variance of the results trained by all available subjects at each speed. After fused according to Eq. (26), the average RMSE of the model trained by all available subjects and the model trained by LOSO cross-validation was 0.026 m/s and 0.042 m/s respectively, and the R-ratio was decreased to 1.62. This indicated higher accuracy and better generalization capability than deep learning methods. However, the fusion parameters for the deep learning method was determined by the training results which was unsuitable for the unseen subjects. To solve this issue, the penalty term was introduced and the fusion strategy was described by Eqs. (27)–(29). According to Eq. (28), when the estimation of the LSTM network model differs a lot from the bias-corrected model-based estimation, the result was unreliable, the weight of the LSTM model would be reduced. The overall RMSE of the fusion model trained by all available subjects was 0.023 m/s after the penalty term was introduced, which was slightly improved than the deep learning method. However, the overall RMSE of the fusion model trained by LOSO cross-validation was decreased to 0.036 m/s which was greatly improved than the deep learning methods and the R-ratio was 1.56 which indicated that the generalization capability was generally high. As was plotted in Fig. 7, the estimated walking speed of the fusion strategy was generally better than deep learning methods. The linear regression and the determination coefficient of each method were shown in Fig. 8, and the function of the best fit line was presented in Table 2. The intercept indicated the systematic error and the slope indicated how whether the estimation is close to unbiased. The overall RMSE and R-ratio of each method were plotted in Fig. 9.

3.4. Time complexity analysis

The most time-consuming procedures of the proposed methods are the training data collection and neural-network training. However, both of these two procedures are performed offline, which means they do not consume any time during the online prediction phase. The proposed method was implemented in Matlab R2019b and performed on a personal computer equipped with an Intel Core i5-7300HQ CPU at 3.50 GHz and 16 GB of DDR4 RAM. Table 3 reported the training and test time of the LSTM network. For 1517 strides dataset, the time consumption of training is 48 min 50 s. From our test, the running time of prediction was less than 2.6 ms (3.884/1517) for each stride.

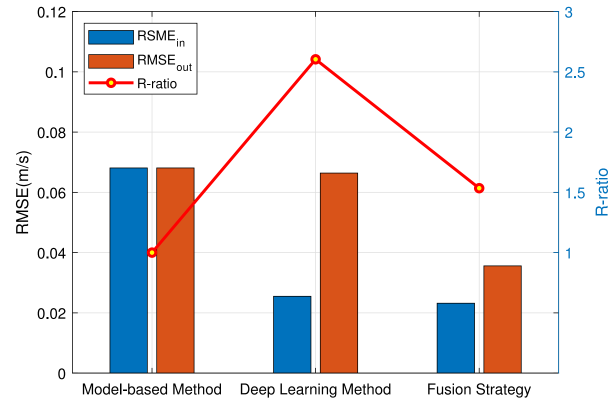

Fig. 9. The RMSE and R-ratio of all the methods.

4. Conclusion

This paper firstly applied LSTM network to train the walking speed estimation model for its ability to consider high-level temporal correlation of multi-dimensional kinematic parameters during one stride. Results showed that the overall RMSE of the model trained by all the available subjects was 0.026 m/s. However, when the LOSO cross-validation was applied to the deep learning method, the RMSE increased to 0.066 m/s and the ratio with respect to the fully trained model was 2.54, which indicated a low generalization capability. For the purpose to realize high accuracy and good generalization capability stride-by-stride walking speed estimation using shank-mounted IMU, a novel fusion strategy based on Bayesian Inference was proposed in this paper. The fusion strategy considers the results derived from different methods using the same raw data as different abstract sensors. After the biases and variances of the abstract sensors were predefined based on the training dataset, the estimation can be easily performed online. Results showed that the overall RMSE of reduced to 0.023 m/s and 0.036 m/s for the fully trained model and LOSO cross-validation respectively and the R-ratio was reduced to 1.56, which indicated that the fusion strategy could achieve high accuracy as well as good generalization capability walking speed estimation.

Although the current study realized the fusion strategy using LSTM model and refined pendulum gait model, the fusion strategy can be expanded to other methods as well. If multi-methods were applied, the fusion strategy can still work by merely changing the weighted index according to the variances of each method. Further studies may fuse the direct integration methods as well and may achieve better performance. In addition, the calculation of the proposed strategy is simple enough to be performed online by embedding devices, which is essential for the applications needing real-time control based on the estimated walking speed, such as the self-paced treadmill. It should also be noted that the proposed strategy can be expanded to other applications where the generalization capability is as important as the accuracy, especially for human physical activity analysis.

CRediT authorship contribution statement

Qian Yuyang: Conceptualization, Methodology, Software, Data curation, Writing - original draft. **Yang Kaiming:** Validation, Writing - review & editing. **Zhu Yu:** Supervision, Project administration. **Wang Wei:** Investigation, Resources. **Wan Chenhui:** Formal analysis, Visualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] T.P. Andriacchi, J.A. Ogle, J.O. Galante, Walking speed as a basis for normal and abnormal gait measurements, J. Biomech. 10 (4) (1977) 261–268, http://dx.doi.org/10.1016/0021-9290(77)90049-5.

[2] H.P. Vonschroeder, R.D. Coutts, P.D. Lyden, E. Billings, V.L. Nickel, GAIT Parameters following stroke - a practical assessment, J. Rehab. Res. Develop. 32 (1) (1995) 25–31.

[3] J.B.J. Bussmann, K.M. Culhane, H.L.D. Horemans, G.M. Lyons, H.J. Stam, Validity of the prosthetic activity monitor to assess the duration and spatio-temporal characteristics of prosthetic walking, Ieee Trans. Neural Syst. Rehab. Eng. 12 (4) (2004) 379–386, http://dx.doi.org/10.1109/tnsre.2004.840495.

[4] S. Godha, G. Lachapelle, Foot mounted inertial system for pedestrian navigation, Meas. Sci. Technol. 19 (7) (2008) http://dx.doi.org/10.1088/0957-0233/19/7/075202.

[5] J.L. Souman, P.R. Giordano, I. Frissen, A. De Luca, M.O. Ernst, Making virtual walking real: Perceptual evaluation of a new treadmill control algorithm, ACM Trans. Appl. Perception 7 (2) (2010) http://dx.doi.org/10.1145/1670671.1670675.

[6] L.H. Sloot, M.M. van der Krogt, J. Harlaar, Self-paced versus fixed speed treadmill walking, Gait Posture 39 (1) (2014) 478–484, http://dx.doi.org/10.1016/j.gaitpost.2013.08.022.

[7] L.H. Sloot, J. Harlaar, M.M. van der Krogt, Self-paced versus fixed speed walking and the effect of virtual reality in children with cerebral palsy, Gait Posture 42 (4) (2015) 498–504, http://dx.doi.org/10.1016/j.gaitpost.2015.08.003.

[8] M. Al-Amri, H. Al Balushi, A. Mashabi, Intra-rater repeatability of gait parameters in healthy adults during self-paced treadmill-based virtual reality walking, Comput. Methods Biomech. Biomed. Eng. 20 (16) (2017) 1669–1677, http://dx.doi.org/10.1080/10255842.2017.1404994.

[9] Y. Qian, K. Yang, Y. Zhu, W. Wang, C. Wan, Local dynamic stability of self-paced treadmill walking versus fixed-speed treadmill walking, J. Biomech. Eng. 142 (4) (2020) http://dx.doi.org/10.1115/1.4045595.

[10] J. Yoon, H.-S. Park, D.L. Damiano, A novel walking speed estimation scheme and its application to treadmill control for gait rehabilitation, J. Neuroeng. Rehab. 9 (2012) http://dx.doi.org/10.1186/1743-0003-9-62.

[11] S. Yang, Q. Li, Inertial sensor-based methods in walking speed estimation: A systematic review, Sensors 12 (5) (2012) 6102–6116, http://dx.doi.org/10.3390/s120506102.

[12] A.M. Sabatini, C. Martelloni, S. Scapellato, F. Cavallo, Assessment of walking features from foot inertial sensing, IEEE Trans. Biomed. Eng. 52 (3) (2005) 486–494, http://dx.doi.org/10.1109/tbme.2004.840727.

[13] A. Mannini, A.M. Sabatini, Walking speed estimation using foot-mounted inertial sensors: Comparing machine learning and strap-down integration methods, Med. Eng. Phys. 36 (10) (2014) 1312–1321, http://dx.doi.org/10.1016/j.medengphy.2014.07.022.

[14] K. Brzostowski, Novel approach to human walking speed enhancement based on drift estimation, Biomed. Signal Process. Control 42 (2018) 18–29, http://dx.doi.org/10.1016/j.bspc.2018.01.002.

[15] C. Shanshan, C.L. Cunningham, J. Lach, B.C. Bennett, Extracting spatio-temporal information from inertial body sensor networks for gait speed estimation, in: 2011 8th International Conference on Body Sensor Networks, 2011, pp. 71–76, http://dx.doi.org/10.1109/bsn.2011.40.

[16] K. Aminian, B. Najafi, C. Bula, P.F. Leyvraz, P. Robert, Spatio-temporal parameters of gait measured by an ambulatory system using miniature gyroscopes, J. Biomech. 35 (5) (2002) 689–699, http://dx.doi.org/10.1016/s0021-9290(02)00008-8.

[17] J.-S. Hu, K.-C. Sun, C.-Y. Cheng, A kinematic human-walking model for the normal-gait-speed estimation using tri-axial acceleration signals at waist location, IEEE Trans. Biomed. Eng. 60 (8) (2013) 2271–2279, http://dx.doi.org/10.1109/tbme.2013.2252345.

[18] W.-S. Yeoh, I. Pek, Y.-H. Yong, X. Chen, A.B. Waluyo, Ieee, Ambulatory monitoring of human posture and walking speed using wearable accelerometer sensors, in: 2008 30th Annual International Conference of the Ieee Engineering in Medicine and Biology Society, Vols. 1-8, in: IEEE Engineering in Medicine and Biology Society Conference Proceedings, 2008, pp. 5184–5187, http://dx.doi.org/10.1109/iembs.2008.4650382.

[19] R.S. McGinnis, N. Mahadevan, Y. Moon, K. Seagers, N. Sheth, J. Wright, S. DiCristofaro, I. Silva, E. Jortberg, M. Ceruolo, J.A. Pindado, J. Sosnoff, R. Ghaffari, S. Patel, A machine learning approach for gait speed estimation using skin-mounted wearable sensors: From healthy controls to individuals with multiple sclerosis, Plos One 12 (6) (2017) http://dx.doi.org/10.1371/journal.pone.0178366.

[20] A. Shrestha, W. Myounggyu, Deepwalking: enabling smartphone-based walking speed estimation using deep learning, in: 2018 IEEE Global Communications Conference, 2018, p. 6, http://dx.doi.org/10.1109/glocom.2018.8647857.

[21] A. Mannini, S.S. Intille, M. Rosenberger, A.M. Sabatini, W. Haskell, Activity recognition using a single accelerometer placed at the wrist or ankle, Med. Sci. Sports Exerc. 45 (11) (2013) 2193–2203, http://dx.doi.org/10.1249/MSS.0b013e31829736d6.

[22] S. Geman, E. Bienenstock, R. Doursat, Neural networks and the bias variance dilemma, Neural Comput. 4 (1) (1992) 1–58, http://dx.doi.org/10.1162/neco.1992.4.1.1.

[23] J.I. Olszewska., Designing transparent and autonomous intelligent vision systems, in: Proceedings of the 11th International Conference on Agents and Artificial Intelligence - Volume 2: ICAART,, INSTICC. SciTePress, 2019, pp. 850–856, http://dx.doi.org/10.5220/0007585208500856.