Personal-specific gait recognition based on latent orthogonal feature space

Quan Zhou1 ε | Jianhua Shan1 | Bin Fang2 ε | Shixin Zhang1 | Fuchun Sun2 | Wenlong Ding1 | Chengyin Wang1 | Qin Zhang3

1AnHui Province Key Laboratory of Special Heavy Load Robot, Anhui University of Technology, Ma Anshan, China
2Beijing National Research Center for Information Science and Technology, Department of Computer Science and Technology, Tsinghua University, Beijing, China
3Huzhong University of Science and Technology, China

Abstract
Exoskeleton has been applied in the field of medical rehabilitation and assistance. However, there are still some problems in the interaction between human and exoskeleton, such as time delay, the existence of certain constraints on the human body, and the movement in time is hard to follow. A human motion pattern recognition model based on the long short-term memory (LSTM) is proposed, which can recognise the state of the human body. Meanwhile, the orthogonalisation method is integrated to make personal-specific disentangling, and it can effectively improve the generalisation ability of different groups of people, so as to improve the effective follower ability of the exoskeleton. Compared with some other traditional methods, this model has better performance and stronger generalisation ability, which has certain significance in the field of exoskeleton algorithm.

1 | INTRODUCTION

With the arrival of the ageing population and the occurrence of various traffic accidents, more people become mobility inconvenience. How to help these people walk or carry out activities better has become an urgent problem to be solved. Fang et al. present a rotation-invariant dynamical-movement-primitive method for learning interaction skills to effectively transfer human behavior skills to a robot [1]. Exoskeleton, as one of the wearable robots, is a good solution to this problem. In order to improve the comfort and performance of wearing exoskeleton, many scholars have studied the human–exoskeleton interaction. Establishing the mapping relationship between human motion data and behaviour can guide the control of exoskeleton and improve the interaction between human and exoskeleton.

Gait pattern signal is commonly performed via inertial measurement units (IMUs), tactile sensors, vertical ground reaction force [2], surface electromyography, electroencephalograms, and so on [3]. The majority of popular gait datasets employ computer vision technology to improve efficiency, such as the Carnegie Mellon University Motion of Body dataset [4], the Human Identification at a Distance—University of Maryland dataset, the Chinese Academy of Sciences Institute of Automation Gait database [5], and the Osaka University Institute of Science and Industry Research Gait database [6]. Chen et al. proposed a method based on hidden Markov model and employed traditional fusion rules to combine the recognition results at the decision level, which resulted in ideal performance [7]. Omar et al. proposed a gait recognition approach for person re-identification, which takes multitask convolutional neural network models and extracted gait energy images (GEIs) to estimate the angle and recognise the gait, and achieves the SOTA performance [8]. Recurrent neural networks, such as long short-term memory (LSTM), have been widely
used in the areas of time series analysis and natural language processing. The cyclic nature of the human gait has previously prevented the use of these networks [9]. Chalvatzi et al. presented a novel LSTM and non-wearable sensor based on the human gait stability prediction framework, which can provide robust predictions of the human stability state [10]. These datasets and methods play an important role in analysing the human gait pattern. However, for the gait analysis algorithm applied to the exoskeleton, video-based data and methods are not applicable.

To achieve the gait recognition algorithm applied to exoskeleton, the data used are acceleration and gyroscopic data of IMU. These data are essentially time series signals. For inertial data, Fang et al. presented a novel data glove for gestures capturing and recognition based on inertial and magnetic measurement units (IMUs), which are made up of three-axis gyroscopes, three-axis accelerometers and three-axis magnetometers. Meanwhile, they used extreme learning machine (ELM) for gesture recognition and achieve an ideal result [11]. Meanwhile, many methods have been proposed for gait analysis and recognition based on inertial data. In the past, plenty of researchers used the traditional manual feature extraction method to extract and analyse the gait features. However, most of those approaches heavily rely on heuristic hand-crafted feature extraction methods, which dramatically hinder their generalisation performance [12]. Shi et al. used IMU and random forest method to recognise five kinds of gait and achieved good results [13].

With the rapid development of deep learning, it makes feature extraction possible to perform automatic high-level feature extraction thus achieves promising performance in this area. Fang et al. proposed the gait neural network (GNN) method based on temporal convolutional neural network (TCN), which can effectively recognises and predicts the human lower limb movement behaviour, and have achieved better results than the traditional methods [14]. Andrey et al. propose using convolutional neural networks (CNN) for local feature extraction together with simple statistical features that preserve information about the global form of time series, and the results show that the proposed model demonstrates state-of-the-art performance while requiring low computational cost and no manual feature engineering [15]. A deep convolutional neural network (convnet) also proposed to perform efficient and effective human activity recognition using smartphone sensors by exploiting the inherent characteristics of activities and 1D time-series signals [16].

However, there are still some problems in exoskeleton. For example, for the lack of human-computer coupling, it can not give a good feedback to human motion, which leads to a certain delay and restricts human movement to a certain extent. Moreover, at present, the generalisation ability of exoskeleton needs to be improved. How to effectively improve the matching degree of various groups of people, instead of just targeting specific wearers, has become a problem.

In practical application, it is difficult for us to collect gait features of all groups of people to train the model. How to improve the generalisation ability of the model on the recognition of lower limb motion behaviour on a small number of samples is of great significance.

We proposed a model based on LSTM and CNN-based person-specific feature disentangling module, which can reduce the influence of human specificity on the model, so as to improve the generalisation ability and better realise the action recognition of different groups. Compared with other methods, we find that the generalisation performance of the model has been significantly improved.

2 | RELATED WORK

The human gait prediction by IMU, sole pressure, myoelectric and EEG signals has also been studied. The HuGaDB dataset of the University Higher School of Economics contains detailed kinematic data for analysing human gait and activity recognition. It is first providing human gait data from inertial sensors and containing segmented annotations for studying the transition between different activities. The data is obtained from 18 participants and totally recording about 10 h [17]. The Robotics Laboratory in Seoul National University collected the data of six gaits through the direct driving of the exoskeleton on the body, such as walking on the flat, walking up the ramp, walking down the ramp, walking upstairs, walking downstairs, and squatting.

The gait data obtained by inertial sensor is a kind of time series data in essence, so the LSTM model can achieve good results. Some researchers’ works also proves this view. Romero et al. has report the application of LSTM for the modeling of gait synchronisation of legs using a basic configuration of off-the-shelf IMU. The model can be transferred to robotised prostheses and assistive robotics devices in order to achieve quick stabilisation and robust transfer of control algorithms to new users [18].

Recently, in the study of human facial action unit detection, Niu et al. have proposed a method based on local relationship learning with person-specific shape regularisation [19], which can effectively extract the pattern features of the same expression, reduce the influence of specific person, and improve the generalisation ability of the model. Niu et al. also proposed a cross-verified feature disentangling strategy to disentangle the physiological features with non-physiological representations which can improve the robust ability of physiological measurement [20].

3 | DATASET

3.1 | Motion capture system

We collect the inertial data of human motion, including acceleration and gyroscope data by IMU. The whole motion capture system have six inertial sensors that located the lower limb. The locations of these sensors are as shown in Figure 1.

The static accuracy of the system is pitch angle, roll angle accuracy is ±1° and heading angle accuracy is ±2°. The maximum measurement range of the system is ±2000 DPS for
angular velocity and ±16 g for acceleration. All sensors are fixed in the designated place by special bandage, and the data is transmitted through hub.

In this study, we collected data on five gait patterns including walking, up and down stairs, slope up and down for nine participants. The participants were all students, both male and female.

As shown in Figure 2, we selected the left thigh y axis acceleration data of walking pattern of four people who participated in the data collection, and we can find that although there are some slight differences, the overall gait presents a kind of similarity. If we can extract useful features, we can achieve better recognition performance.

\[ x_{\text{norm}} = \frac{x_i - \text{mean}(x_i)}{\text{std}(x_i)} \]  

where \( x_i \) is the \( i \)-th channel gait data (acceleration or quaternion). After normalising the data, the model can converge better.

Considering the possible similarity of gait data at a certain time point under different motion modes, we do not take the data of a single time step as the input of the model, but select a period of time data as the input. The sequence length of gait data is set to 10.

We collected data on five gait pattern, including walking, going up and down stairs, up and down slopes. The label dictionary is as follows: \{walking: 0, upstairs: 1, downstairs: 2, slopeup: 3, slopedown: 4\}

4 | METHODS

4.1 | Whole model

In this paper, we use LSTM as the stem module to extract features related to lower limb pattern recognition. At the same time, we use a parallel CNN module to extract features related to personal information. Our parallel module mainly includes a
convolutional layer, an average pooling layer and a full connection layer to classify personal information.

The whole model structure is shown in Figure 3. Through this structure, we want to achieve that the features extracted by CNN module are more related to personal information than gait, and the features extracted by LSTM module are more related to gait action than to personal information. In order to achieve this goal, we use the idea of cosine similarity, calculate the inner product of the state vector of the last hidden layer of LSTM and the vector of the last feature extracted layer of CNN, and get the result as separation loss. By strengthening the ability of the two modules to extract different information, the stem module can have stronger feature extraction ability to the general gait data information, so as to improve the generalisation ability of gait recognition.

4.2 LSTM (baseline)

The gait data acquired by motion capture system is time series data in essence, so LSTM model is used for further processing and feature extraction. LSTM is widely used in speech recognition, sequence data processing. Compared with the traditional recurrent neural network (RNN), LSTM can remember the information in the time series signal in the past by gating mechanism.

As shown in Figure 4, the structure of LSTM (baseline) is provided.

The calculate formula of LSTM is as follows:

\[ i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i), \]
\[ f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f), \]
\[ o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o), \]
\[ \tilde{c}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \]
\[ c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \]  

(2)

where \( i_t, f_t, o_t \) are input gate, forget gate, output gate probability, respectively.

Through LSTM-based stem model, we can get the information of gait data on time series. At the same time, the final hidden state will be output as the general gait feature, and the final full connection layer will complete the classification task of the action. In this way, the final hidden state vectors have a strong correlation with general gait information.

4.3 Parallel module

As a deep learning model, CNN can effectively extract the characteristics of data. The data we input into the CNN model is in the form (sequence length * sensor channels). We used the feature extraction capability of CNN to extract the feature containing personal information of the matrix and obtained the corresponding feature vector. Finally, the feature vector is feed to the full connection layer to output the classification result of personal information (ID). In this way, the extracted feature vectors have a strong correlation with personal information (ID).

As shown in Figure 5, the structure of parallel CNN module is provided. We use a convolution layer and an pooling
layer to extract the relevant features. In addition, considering the noise in gait data, we use the average pooling layer instead of the maximum pooling layer, which can reduce the abnormal results caused by noise.

The calculation formula for the convolution layer is:

\[ x_j^n = f \left( \sum_{i \in M_j} x_i^{n-1} \ast k_{ij}^n + b_j^n \right) \]  \hspace{1cm} (3)
where \( x_i^n \) is the output feature map, \( x_i^{n-1} \) is the input feature map, \( M_i \) is the selected area in the \( n-1 \) layer, \( k_{ij}^n \) is weight parameter, \( b_j^n \) is bias, \( f \) is activation function.

The calculation formula for the fully connected layer is:

\[
y_j = f \left( \sum_{i=1}^{N} x_i^n w_{ij} + b_j \right)
\]

(4)

where \( x \) is the input layer, \( N \) is number of input layer nodes, \( w_{ij} \) is the weight between the links \( x_i \) and \( y_j \), \( b_j \) is the bias. \( f \) is activation function.

Finally, the full connection layer outputs the personal information (ID) of different data providers, so that the features extracted by CNN have strong correlation with personal information and we can do the inner product with another set of feature vectors from LSTM to achieve feature disentangling.

### 4.4 Loss function

In the training process of the model, we will get the feature state vector in the middle output of the model, and after dotting the two, we will get separation loss as one part of total loss. Through this loss function, we can separate the general gait characteristics from the individual characteristics as far as possible, so as to improve the generalisation ability of the model.

In addition, we use the cross-entropy loss as the recognition loss function.

\[
L_{sep} = f_{CNN} \odot f_{LSTM},
L_{total} = L_{cls} + \alpha L_{sep} + \beta L_{id}.
\]

(5)

where \( \alpha \) and \( \beta \) are the balance hyper-parameters, \( L_{cls} \) is the loss of gait pattern recognition, \( L_{sep} \) is the inner product of CNN and LSTM feature vectors and \( L_{id} \) is the loss of person recognition.

In the process of model iteration, the total loss will continue to decrease, which means that \( L_{sep} \) is also decreasing, which indicates that the similarity of feature vectors extracted from the two modules is decreasing, that is, the feature vectors extracted by LSTM will have stronger correlation with general gait features, while the feature vectors extracted by CNN module will be more relevant to personal information.

In the validation of generalisation ability, only the recognition performance of gait pattern is considered, and the classification loss of personal information \( L_{id} \) can be regarded as the regularisation term of the weight of LSTM model.

### 4.5 Performance measures

Although our sample size is approximately balanced, considering the performance of a more comprehensive evaluation model, it is not enough to only use the accuracy rate. Therefore, we use a number of evaluation indicators to prove the superiority of our model.

The indicators of performance measure are as follows:

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}
\]

\[
\text{Recall} = \frac{TP}{TP + FN}
\]

\[
\text{Precision} = \frac{TP}{TP + FP}
\]

\[
F1 \text{ Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}
\]

where true positive (TP), which is samples of positive that are correctly recognised by the model. False positive (FP), which is samples of negative that are recognised as positive by the model. True negative (TN), which is samples of negative that are correctly recognised by the model. False negative (FN), which is samples of positive that are recognised as negative by the model.

### 5 EXPERIMENTS AND RESULTS

In order to verify the generalisation ability of the model, we carried out tests on small-sample and larger-sample respectively.

#### 5.1 Experimental approach

In this study, the performance of our model was evaluated on a human gait dataset obtained using an inertial-based wearable motion capture device. The model was trained by an Adam optimizer with a learning rate of 0.001 and at 50 epochs, divided by 10. The maximum epoch and batch size were 100 and 64, respectively. The dropout rate of all dropout layers was set to 0.3. The model was implemented by PyTorch and trained and tested on a computer with an Intel Core i7-8750H, two 8 GB memory chips (DDR4), and a GPU (GeForce GTX 1060 6G).

#### 5.2 Experiment on small-sample

In order to better evaluate the performance of the model, we selected the gait data of four collectors for validation. The gait data of three data providers is used as training set, and the other data not participating in training is used as test set to verify the generalisation ability of the model.

In order to prove the effectiveness of the method, we use some other methods to compare the classification effect on the same dataset, including CNN, support vector machine, back propagation and LSTM. In addition, the LSTM model structure is the same as stem module in our personal-specific disentangle model, and in CNN model, the gait data is transformed into a matrix whose length is the sequence length and the width is the sensor channels as input.

We collected 2574 gait data that did not participate in the training, which included five types of movements. And on
these data, we verify the performance of all the models. As shown in the Table 1, we provide the verification performance of different models on five types of actions. We can find that our model shows better performance than other models and achieves better results in gait recognition for different groups of people.

As shown in Figures 6, we provide the confusion matrix of our model and we can find that our model can better make gait pattern recognition. At the same time, we found that walking up and down stairs and uphill is easy to misjudge.

### 5.3 | Experiment on larger-sample

Considering the characteristics of the data we collected and the gait data of the whole population in practical application, we conducted further experiments. We used the data of five gait movements of five people as the training set, and the data of five movements of the other four people as the test set and the size of test set is 47,485.

Similarly, we compared the performance of other models, and found that when the samples of training data increased, the performance of all methods were improved, but our model still has some advantages. Considering the increase of the number of samples, more gait pattern features will be included, so it is inevitable to improve the effect of each model, but when the number of samples is small, our method is still very meaningful. At the same time, it should be added that the parameters $\alpha$ and $\beta$ may need to be adjusted to obtain the best effect for different data samples.

As shown in Table 2, we provided the comparison result of larger-sample. We can find that when the number of samples increases, the accuracy of action recognition based on misclassification in the case of small samples is also improved.

| Action      | Method           | Precision | Recall | F1-Score | Support | Average Accuracy |
|-------------|------------------|-----------|--------|---------|---------|-----------------|
| Walking     | Ours ($\alpha = 0.4, \beta = 0.9$) | [1.00] | 0.97   | 0.98    | 8740    | [0.87]          |
| Upstairs    |                  | [0.84]   | 0.79   | [0.82]  | 8488    |                 |
| Downstairs  |                  | [0.77]   | 0.82   | 0.79    | 9047    |                 |
| Slope up    |                  | 0.85     | [0.90] | [0.87]  | 10,268  |                 |
| Slope down  |                  | 0.91     | [0.88] | [0.89]  | 10,942  |                 |
| Walking     | LSTM (baseline)  | [0.99]   | [1.00] | [0.99]  | 8740    | 0.82            |
| Upstairs    |                  | 0.66     | 0.68   | 0.67    | 8488    |                 |
| Downstairs  |                  | 0.64     | 0.84   | 0.73    | 9047    |                 |
| Slope up    |                  | [0.89]   | 0.83   | 0.86    | 10,268  |                 |
| Slope down  |                  | [0.96]   | 0.74   | 0.84    | 10,942  |                 |
| Walking     | CNN              | 0.97     | 0.97   | 0.97    | 8740    | [0.86]          |
| Upstairs    |                  | [0.74]   | [0.82] | [0.78]  | 8488    |                 |
| Downstairs  |                  | [0.83]   | 0.85   | [0.84]  | 9047    |                 |
| Slope up    |                  | 0.88     | 0.83   | 0.86    | 10,268  |                 |
| Slope down  |                  | 0.89     | [0.85] | [0.87]  | 10,942  |                 |
| Walking     | BP               | 0.98     | [0.99] | 0.98    | 8740    | 0.79            |
| Upstairs    |                  | 0.58     | 0.48   | 0.53    | 8488    |                 |
| Downstairs  |                  | 0.68     | [0.90] | 0.78    | 9047    |                 |
| Slope up    |                  | 0.80     | [0.84] | 0.82    | 10,268  |                 |
| Slope down  |                  | 0.91     | 0.74   | 0.82    | 10,942  |                 |
| Walking     | SVM              | [1.00]   | [1.00] | [1.00]  | 8740    | 0.85            |
| Upstairs    |                  | 0.68     | 0.73   | 0.70    | 8488    |                 |
| Downstairs  |                  | 0.73     | [0.93] | [0.82]  | 9047    |                 |
| Slope up    |                  | [0.98]   | [0.90] | [0.94]  | 10,268  |                 |
| Slope down  |                  | [0.94]   | 0.74   | 0.83    | 10,942  |                 |

Note: The best and second are indicated using brackets and bold, and brackets alone, respectively.

Abbreviations: BP, back propagation; CNN, convolutional neural networks; LSTM, long short-term memory; SVM, support vector machine.
sample. In the small-sample experiment, we can find that our model can better separate the features, which are more related to the gait pattern, so as to improve the generalisation ability of the model in different groups of people. In the larger-sample experiment, when the sample size is large enough, the features extracted by the model may be enough to cover the gait of all kinds of people, and the effect may not be so significant. But even so, considering the complexity of data collection in a real environment, our model is still meaningful. From the results of these two experiments, we can find that our model can effectively extract the general gait features of different populations. Thus, when new untrained data are input, the model can recognise the corresponding actions more accurately. At the same time, through the confusion matrix, we can find that the classification result of all models is not very ideal for the up and down stairs gait, which may be caused due to the similarity between the upper and lower gait, which is also a problem to be solved in the future, that is, how to correctly classify the gait with similarity.

In future work, we can consider more and more complex action categories to identify tasks. At the same time, how to effectively identify actions with certain similarity is also a very meaningful problem.

ACKNOWLEDGEMENT
This work was supported by the Natural Science Foundation of University in Anhui Province (No.KJ2019A0086) and the National Key Research and Development Program of China (Grant No. 2017YFB1302302).

ORCID
Quan Zhou @ https://orcid.org/0000-0003-2959-8985
Bin Fang @ https://orcid.org/0000-0002-9149-7336

REFERENCES
1. Fang, B., et al.: Skill learning for human-robot interaction using wearable device. Tsinghua Sci. Technol. 24(6), 654–662 (2019)
2. Karout, H., et al.: Frequency content analysis of gait — vertical ground reaction force. In: 2015 International Conference on Advances in Biomedical Engineering (ICABME), pp. 53–56. IEEE (2015). https://doi.org/10.1109/ICABME.2015.7323249
3. Nweke, H., et al.: Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: state of the art and research challenges. Expert Syst. Appl. 105 (2018). https://doi.org/10.1016/j.eswa.2018.03.056
4. Gross, R., Shi, J.: The CMU motion of body (MoBo) database, 27 (2001). https://doi.org/10.1515/9781400843251-003
5. Zheng, S., et al.: Robust view transformation model for gait recognition. In: Proceedings - International Conference on Image Processing ICIP. 2073-2076 (2011). https://doi.org/10.1109/ICIP.2011.615889
6. Iwama, H., et al.: The OU-ISIR gait database comprising the large population dataset and performance evaluation of gait recognition. IEEE Trans. Inform. Forens. Sec. 7, 1511–1521 (2012). https://doi.org/10.1109/TIFS.2012.2204253
7. Chen, C., et al.: Factorial HMM and parallel HMM for gait recognition. IEEE Trans. Syst. Man, Cyber., Part C: Appl. Rev. 39, 114–123 (2009). https://doi.org/10.1109/TSMCC.2008.2001716
8. El Harrouss, O., et al.: Gait recognition for person re-identification. J. Supercomput. (2020). https://doi.org/10.1007/s11227-020-03409-5
9. Martinez, J., Black, M., Romero, J.: On human motion prediction using recurrent neural networks. 4674–4683 (2017). https://doi.org/10.1109/CVPR.2017.497

10. Chalvatzaki, G., et al.: On-line human gait stability prediction using LSTMs for the fusion of deep-based pose estimation and LRF-based augmented gait state estimation in an intelligent robotic rollator (2018)

11. Fang, B., et al.: 3D human gesture capturing and recognition by the IMMUBased data glove. Neurocomput. 277 (2017). https://doi.org/10.1016/j.neucom.2017.02.101

12. Wang, J., et al.: Deep learning for sensor-based activity recognition: a survey. Pattern Recognit. Lett. (2017). https://doi.org/10.1016/j.patrec.2018.02.010

13. Shi, L.-F., et al.: Gait recognition via random forests based on wearable inertial measurement unit. J. Ambient Intell. Humaniz. Comput. (2020). https://doi.org/10.1007/s12652-020-01870-x

14. Fang, B., et al.: Gait prediction network developed for human-exoskeleton interaction. Front. Neurorobot. https://doi.org/10.3389/fnbot.2020.00058

15. Andrey, I.: Real-time human activity recognition from accelerometer data using convolutional neural networks. Appl. Soft. Comput. 62 (2017). https://doi.org/10.1016/j.asoc.2017.09.027

16. Romao, C., Cho, S.-B.: Human activity recognition with smartphone sensors using deep learning neural networks. Exp. Syst. Appl. 59 (2016). https://doi.org/10.1016/j.eswa.2016.04.032

17. Chereshevn, R., Kertesz-Farkas, A.: HuGaDB: human gait Database for activity recognition from wearable inertial sensor networks (2018). https://doi.org/10.1007/978-3-319-73013-4_12

18. Romero-Hernandez, P., Asiain, J., Grana, M.: Deep learning prediction of gait based on inertial measurements (2019). https://doi.org/10.1007/978-3-030-19591-5_29

19. Niu, X., et al.: Local relationship learning with person-specific shape regularization for facial action unit detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 11909–11918, 2019. https://doi.org/10.1109/CVPR.2019.01219

20. Niu, X., et al.: Video-based remote physiological measurement via cross-verified feature disentangling (2020)

How to cite this article: Zhou Q, Shan J, Fang B, et al. Personal-specific gait recognition based on latent orthogonal feature space. *Cogn. Comput. Syst.* 2021;3:61–69. https://doi.org/10.1049/ccs2.12007