BIG DATA MINING AND ANALYTICS
ISSN 2096-0654 01/06 pp223–232
Volume 4, Number 4, December 2021
DOI: 10.26599/BDMA.2021.9020006

Multimodal Adaptive Identity-Recognition Algorithm Fused with Gait Perception

Changjie Wang, Zhihua Li*, and Benjamin Sarpong

Abstract: Identity-recognition technologies require assistive equipment, whereas they are poor in recognition accuracy and expensive. To overcome this deficiency, this paper proposes several gait feature identification algorithms. First, in combination with the collected gait information of individuals from triaxial accelerometers on smartphones, the collected information is preprocessed, and multimodal fusion is used with the existing standard datasets to yield a multimodal synthetic dataset; then, with the multimodal characteristics of the collected biological gait information, a Convolutional Neural Network based Gait Recognition (CNN-GR) model and the related scheme for the multimodal features are developed; at last, regarding the proposed CNN-GR model and scheme, a unimodal gait feature identity single-gait feature identification algorithm and a multimodal gait feature fusion identity multimodal gait information algorithm are proposed. Experimental results show that the proposed algorithms perform well in recognition accuracy, the confusion matrix, and the kappa statistic, and they have better recognition scores and robustness than the compared algorithms; thus, the proposed algorithm has prominent promise in practice.

Key words: gait recognition; person identification; deep learning; multimodal feature fusion

1 Introduction

Mobile internet networks are applied in fields such as health care and smart cities. They have not only a communal and ubiquitous nature but a social perception function. Much physical and behavioral biometric information can be collected through mobile smart devices connected to a network. Using the perception function of the mobile internet, a variety of biometric feature information can be extracted, and this can be used to identify individuals[1].

Gait recognition can identify people by their walking style[2]. Compared with recognition technologies based on faces and fingerprints, gait recognition has the advantage of being a noncontact approach, and one's gait is easy to perceive and difficult to hide and disguise. Gait recognition research focuses on computer vision, footwear pressure sensing, and acceleration sensing. Computer vision methods are based on image and video processing. Gait recognition usually relies on surveillance cameras and other sensing devices to collect video information, i.e., a video sequence of gaits is obtained by tracking and monitoring a subject. Gait features are extracted by preprocessing and analysis, and gait recognition is performed using technologies related to image processing[3]. These technologies are relatively expensive, and the data are sensitive to environmental noises, such as the installation angle and location of surveillance cameras, which affect identification accuracy. Preprocessing is complex and requires motion detection, segmentation, and feature extraction on acquired image sequences. This requirement contributes to the complexity of data processing. The results are susceptible to light and background noise disturbance. Hence, there is much space for development in application. Methods based on footwear pressure sensing realize recognition using the collected reaction

• Changjie Wang, Zhihua Li, and Benjamin Sarpong are with Department of Computer Science and Technology, School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China. E-mail: thewang@gmail.com; zhli@jiangnan.edu.cn; 6181910030@stu.jiangnan.edu.cn.
* To whom correspondence should be addressed.
Manuscript received: 2020-10-22; revised: 2021-03-23; accepted: 2021-04-26

forces generated by the test subjects' pressure on sensing pads or sensing boards when standing or walking. They require special sensing boards and have limitations in practice. Acceleration sensing based gait recognition technology is an emerging field. It identifies a person by acceleration signals generated by walking. It has the advantages of gait being difficult to disguise and less susceptible to environmental factors[3]. It extracts gait features while the subject is walking; hence it can achieve continuous identity verification with improved security. It has attracted the attention of many researchers, but its recognition accuracy is not high. In addition, because of the disturbance from factors, such as jitter and white noise, gait acceleration signals must be smoothed and filtered, which increases its cost and complexity. Deep learning-related algorithms have excellent capabilities for feature extraction, are less susceptible to covariates and other factors, and are highly efficient. Deep learning algorithms, as represented by Convolutional Neural Networks (CNNs), are used in mature applications of identity-recognition systems, such as voice and facial recognition[4]. Based on these studies, this paper proposes gait recognition algorithms and methods based on deep learning theory.

A single biometric feature is prone to environmental limitation in multi-scene applications. Therefore, we combine the proposed algorithm with datasets used for real-world data mining to develop synthetic personal gait datasets and design a multi-feature fusion person identification algorithm. The system uses sensors on mobile smartphones and other devices to collect personal gait data, implements a CNN-based scheme to classify and train the gait dataset, and creates a gait recognition model based on deep learning. The distinction in gait recognition results of individuals is used to determine identities. Moreover, a clustering method is added to the person recognition algorithm to perform a priori classification of the gait information of the subject to improve the recognition accuracy. Experimental results show that the proposed model has a relatively low cost and high recognition accuracy.

This paper has seven sections. Section 1 introduces the research background. Section 2 describes the current research work in the field of identity recognition. Section 3 analyzes the gait characteristics. Section 4 presents the proposed gait recognition model based on mobile social perception. The recognition model of the person recognition system combined with gait and multi-feature fusion is introduced in Section 5. Section 6 discusses our experiment, which verifies the validity and fault tolerance of the model. Finally, the method proposed in this paper is summarized in Section 7.

2 Related Work

In 1994, Niyogi and Adelson[5] first proposed the use of gait features for identity-recognition. The image information of gait was used to identify individuals. This method has been widely tracked by researchers and found to have low recognition accuracy. Ailisto et al.[6] studied the gait videos of 122 people and found the recognition rate was only 78%. Gait recognition methods have developed rapidly in recent years because of their low cost and excellent robustness. They can be separated into model-based and model-free methods. Tafazzoli and Safabakhsh[7] created an a priori initial model of human leg and arm movements based on anatomical proportions and constructed the posterior model on articulated parts of the body in motion, using active contour models and Hough transforms. Although model-based gait recognition methods have the advantages of angle and scale invariance and insensitivity to background noise, they pose strict requirements on gait quality, while model-free methods omit a priori model construction and pose laxer requirements on gait quality. Mogan et al.[8] proposed a method based on temporal gradients using temporal characteristics of gait movement, which not only describes the contour shape of the spatial gait movement but implicitly captures the contour changes of gaits on the time axis. Although the two types of gait recognition methods differ in terms of the feature-representing method, they both extract gait features and then perform feature-matching and movement-pattern analysis. However, the recognition accuracy of single-stage features still must be improved.

With the development of the perception function of the mobile internet, the concept of information fusion has become practical. In 1997, Bigün et al.[9] proposed the multimodality concept based on information fusion theory. Using mobile phones and various sensors, they extracted the data features of information collected by sensors and combined these to further ensure robust identification. On the basis of the study of fusion theory of multiple biometric features, Jain et al.[10,11] proposed a biometric identification technique based on the fusion of fingerprints and facial features and proved that the multimodal biometric identification technique outperforms unimodal biometric identification. Baig et al.[12] used a Hamming distance matcher to fuse iris and fingerprint features and selected the best fusion feature among them for more accurate identification of individuals. Li et al.[13] proposed an identity-recognition method combining facial and gait information to improve traditional single-gait recognition methods. Bazin and Nixon[14] proposed a probabilistic fusion method of static and dynamic gait features, which effectively improved the complementarity of features. However, the dynamic gait feature extraction adopts a model-based gait feature extraction method[15] and thus is complicated.

Mantyjarvi et al.[16] attached multiple accelerometers to a subject's body. Although collecting static and dynamic gait trait data is easily accomplished, it requires high equipment investment. To overcome the deficiency of the aforementioned methods, we use the mobile internet perception function for person identification. A smartphone is used as a component node of the perception network, and the triaxial accelerometer built in the smartphone is used to collect the temporal sequences of the subject's gait information. Combining deep learning theory and technology, a model-free gait recognition algorithm that clusters and then recognizes the gait time series is proposed. The proposed algorithm improves the applicability and security of person identification.

3 Gait Feature Analysis

Gait is biometric information that is difficult to imitate and is highly unique and secure. Each person's gait is different. When more features are considered, an individual's gait is unique[1]. Triaxial accelerometers are easy to carry, and they can capture accurate and comprehensive gait information. A triaxial accelerometer in a mobile smart terminal has unique characteristics. For example, vertical and forward accelerations show periodic changes in a person's horizontal walking motions. During the swing phase of walking, the vertical acceleration increases in a positive direction because the center of gravity moves upward with one foot touching the ground. As the person continues to move forward, the center of gravity moves downward, the feet touch the ground, and the acceleration shows a downward trend. Horizontal acceleration decreases during the swing phase and increases when taking a step forward. Studies have shown that acceleration in the direction of gravity reaches a minimum at the initial point when the foot steps forward[7] and reaches a maximum when the entire foot touches the ground. The biometric information on three axes collected by a triaxial accelerometer is essentially multimodal.

Usually, gait information collected by a smartphone's triaxial accelerometer is stored as a vector $(G_x, G_y, G_z)^{\mathrm{T}}$, where G_x, G_y, and G_z are the values recorded by the accelerometer in the horizontal, vertical, and forward directions, respectively. When a mobile smart terminal's accelerometer collects gait information, the vertical sensor is likely to yield offset with the sensor vertical to the ground as the person moves. The actual acceleration values corresponding to the gait data can be expressed by Eq. (1), which are measured by the triaxial accelerometer[1,7],

$$\begin{cases} G'_x = G_x \sin\beta - G_y \cos\beta + G_z \cos\beta, \\ G'_y = G_y \sin\beta \sin\alpha - G_z \sin\beta \cos\alpha + G_z \cos\beta \sin\alpha, \\ G'_z = G_y \sin\beta \cos\alpha + G_z \sin\beta \sin\alpha - G_z \cos\beta \cos\alpha \end{cases} \tag{1}$$

where α is the angle between the vertical-direction sensor and the direction of gravitational acceleration g, and β is the angle between the direction of gravitational acceleration projected on the horizontal plane and the person's movement direction.

In this way, $\left(G'_x, G'_y, G'_z\right)$ consists of the actual acceleration corresponding to a person's gait information. Because this information contains disturbances, such as jitter and white noise, we perform smoothing and filtering using the methods in Refs. [17,18].

Figure 1 is a schematic diagram of the result of smoothing and filtering a person's gait information. In Fig. 1, the horizontal axis represents the measurement time, and the vertical axis represents the acceleration values recorded by the accelerometer in the x-, y-, and z- dimensions.

Accelerometer-based gait identification methods mainly use the processing methods of periodic extraction or data segmentation. Periodic extraction is simple and direct, and requires no data conversion, but it is prone to error because the period cannot be accurately determined. Therefore, we use a data segmentation method to process gait samples. In addition, because the human body's gait triaxial acceleration signal exhibits periodicity with peaks and troughs, a person's peak-to-peak distance is consistent[6,7]. Troughs are correlated to the human walking speed and the variation in the acceleration signal caused by the weightlessness of the body during walking. When the walking speed increases, so do the variations

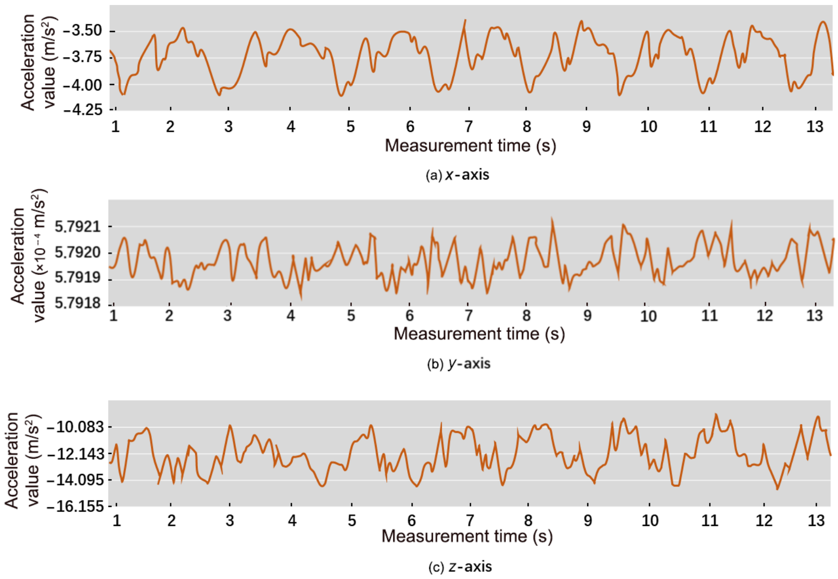


Fig. 1 Line chart of individual gait information demo.

in the waveform, and similarly, the variations decrease with decreased walking speed. We consider the peak interval, standard deviation, and average composition of accelerations of a triaxial acceleration curve as gait trait values and consider four consecutive peak intervals as a sample. To completely remove noise, gait trait values are normalized by the standard deviation during data preprocessing,

$$F' = (F - F_{\text{mean}})/F_s \tag{2}$$

where F' is the final gait trait obtained after preprocessing, F is the original gait trait value, and F_{mean} and F_s are the average and the standard deviation of the gait trait value, respectively.

After normalizing the gait data through Eq. (2), the acceleration values are discretized, and the discretized gait information data are obtained.

4 CNN-Based Gait Recognition Model Training

A CNN[19] is a deep pooling infrastructure consisting of multiple convolutional layers. It includes input, output, fully connected, pooling, and convolutional layers. Because gait behaviors of an individual in different scenarios and motion status have different feature representations, to extract and train gait information, we design and develop a CNN-based Gait Recognition (CNN-GR) model.

During the training stage, the input includes three-dimensional discretized gait information on the x-, y-, and z-axis. The output layer uses the softmax function for normalization to constrain the acceleration signals between 0 and 1. In gait feature extraction, the convolution kernel sequentially performs convolution operations on different positions of the gait information with a specified stride (e.g., four consecutive peak intervals). Because its weight generally does not change during convolution, a convolution kernel can only extract local features in the gait information. To extract gait information features more comprehensively, we use multiple convolution kernels and set up different convolution kernel parameters to process the gait information. In the convolution process, to effectively extract the feature on the edge data in the gait information, zero padding is applied to the gait data edge. After the gait data are convolved, the corresponding feature maps of specific sizes and dimensions can be obtained. The dimensionality of the gait feature maps can be calculated,

$$G_{\text{output}} = \frac{G_{\text{input}} + 2 \times \text{padding} - \text{kernel\_size}}{\text{stride}} + 1 \tag{3}$$

where G_{input} is the dimensionality of the input gait information, "kernel_size" is the size of the convolution kernel, "stride" is the stride of the convolution kernel, "padding" is the number of added padding values at the edge of the input gait data, and G_{output} is the dimension of the output gait feature after one convolution operation.

The CNN-GR model includes four convolutional layers: Conv1, Conv2, Conv3, and Conv4. The first convolutional layer, i.e., Conv1, uses a kernel to filter the input and extracts the shallow-layer features of the gait data. It processes each input vector separately. Its kernel stride is 1. To further extract the deep-layer features of the gait data, Conv2 uses a convolution kernel to calculate the gait information features output by Conv1. The kernel stride of Conv2 is 2. Deep convolutions help to fully extract the feature of gait information. The third and fourth layers, Conv3 and Conv4, have convolution kernels and convolution kernel strides of 1.

Each of the four convolutional layers performs feature extraction on gait feature data that are output by their previous layers. However, a convolutional layer only performs linear transformations on features, and the multilayer convolution operation superimposes the linear transformation of the feature data, with the result that the final feature model is a complex linear model with weak representation ability. Therefore, an activation function must be used to introduce nonlinear operations to the CNN. The CNN-GR model uses a Rectified Linear Unit (ReLU) function as the activation function of the feature data for post-convolution operations. The conceptual diagram of the ReLU function is shown in Fig. 2, and the function can be expressed as

$$f(x) = \max(0, x) \tag{4}$$

The ReLU function is piecewise linear. When the input value is less than 0, its output value is negative; and when the input value is positive, its output value remains unchanged. Therefore, the ReLU function has the characteristic of unilateral suppression, which can introduce nonlinear factors to the CNN-GR model and better mine objective features and fit training data.

Gait information features can be effectively extracted using the four-layer convolution operation and ReLU function. However, as the number of convolutional layers increases, the dimensions of the obtained gait features increase rapidly, as does redundant data. To reduce the dimensions of the features and remove redundant data, i.e., to prevent overfitting, pooling layers are periodically placed between convolutional layers to perform pooling operations on gait feature data. Similar to a convolution operation, a pooling operation has two parameters: the size of the pooling kernel and the pooling stride. After the input gait feature data are pooled, their dimensions can be calculated,

$$p_{\text{output}} = \frac{p_{\text{input}} - \text{kernel\_size}}{\text{stride}} + 1 \tag{5}$$

where p_{input} is the size of the gait feature data that are input to the pooling layer, and kernel_size and stride are identical to those in Eq. (3).

To pool the data, the CNN-GR model uses max pooling, i.e., it selects the maximum value in a feature map region scanned by the pooling window as the result of the pooling operation. Figure 3 is an example in which the right subgraph is the result of maximizing the pooling operation. Obviously, the pooling operation on the feature data can effectively reduce feature dimensionality, the risk of overfitting, and the computational and time complexity of the entire model. The CNN-GR model uses a 2×2 pooling kernel and a stride of 2.

The fully connected layers map the gait feature data extracted by the convolution and pooling operations to the one-dimensional feature vector, which helps to solve the classification or regression problem. The CNN-GR model has two fully connected layers, Dense1 and Dense2. For Dense1, the size of the convolution kernel equals the feature dimension output by Conv4, the stride is 1, and there is no padding being omitted. For Dense2, the fully connected operation is the convolution operation with a 1×1 kernel and a stride of 1. The feature data from all nodes of the fully connected layers are classified by a softmax function,

$$y(x_k) = \frac{\exp(x_k)}{\sum_{i=1}^{M} (\exp(x_i))} \tag{6}$$

where M is the number of feature data points input on x-, y-, and z-axis, and x_k is the k-th input feature data point.

The CNN-GR model, as shown in Fig. 4, is used to

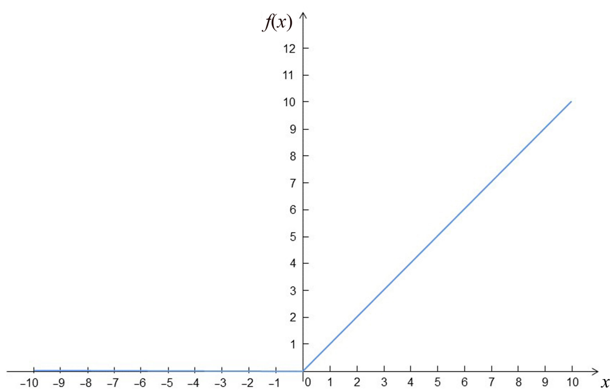


Fig. 2 ReLU function graph.

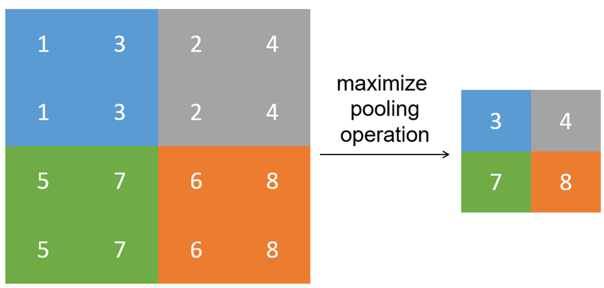


Fig. 3 Max pooling demo.

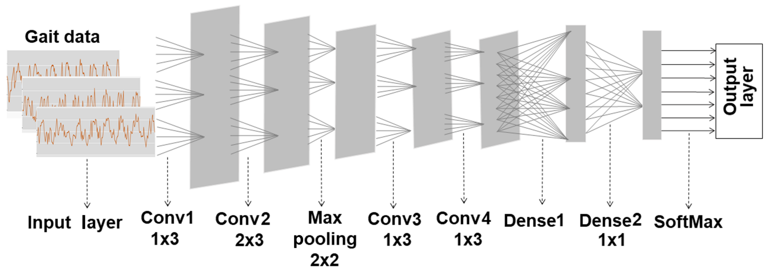


Fig. 4 CNN-GR infrastructure.

train individual gait behavior datasets, and it uses the mutual exclusion of features and the softmax function to complete the recognition and classification of gait behavior.

5 CNN-GR-Based Identification Method

5.1 Unimodal identification algorithm

5.1.1 Unimodal identification scheme

On the basis of the CNN-GR model, we propose a unimodal gait feature recognition scheme and an identity-recognition algorithm. The method uses a smartphone with a triaxial accelerometer to collect gait information in different states and sends it to a backend server, which trains and tests the collected gait dataset on the x-, y-, and z-axes and extracts and preprocesses the gait feature information, which is used to train the CNN-GR model for the gait information of this scheme. Figure 5 shows the flowchart of this scheme.

5.1.2 Unimodal gait-identification algorithm

In combination with the aforementioned identification scheme based on the unimodal gait features, we propose a Single-Gait Feature Identification (SGFI) algorithm, as described by Algorithm 1.

On the basis of the CNN-GR model, the SGFI

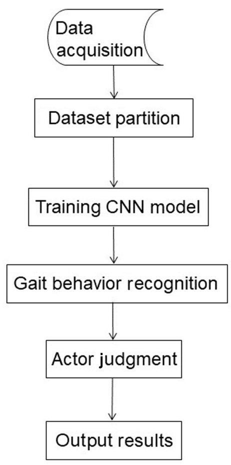


Fig. 5 Unimodal gait feature recognition model.

Algorithm 1 SGFI

Require: Gait training set X
Ensure: Recognition result
Initialize MaxEpoch; // Initialize iterations
Initialization weights; // Initialize parameters
for epoch = 1 to MaxEpoch **do**
X_Feture = Conv(X); //Extract features on convolutional layer
Max_Pool(X_Feture); //Reduce feature vector dimension on pooling layer
Dense(X_Feture); //Connect the feature vector to the fully connected layer
SoftMax(X_Feture); //Calculate the actual output of the sample
Update all weights; //Update the ownership value in the neural network
end for
return result //Come to conclusion

algorithm extracts and classifies the features from the monomodal gait feature data and conducts biometric identification based on uniaxial acceleration information (e.g., monomodal features). Therefore, its time complexity is consistent with that of the CNN model, $O\left(\sum_{l=1}^{D} M_l^2 \cdot K_l^2 \cdot C_{l-1} \cdot C_l\right)$, where D is the number of convolutional layers of the CNN model or the depth of the convolutional network; l is the l-th convolutional layer; M is the edge length of the feature maps output by each convolution kernel; K is the edge length of each convolution kernel; and C is the number of channels of the convolution kernels.

5.2 Multimodal adaptive identification method

5.2.1 Multimodal gait feature fusion

In Section 5.1, the CNN was trained and classified by extracting triaxial acceleration features as gait feature values, and a unimodal gait identification method was based on the proposed CNN-GR model. However, in practice, the walking style of a person can be divided into different types based on activities, such as standing, jogging, walking, and sitting. The characteristics of a person's triaxial acceleration curves differ by type of gait sequence. Hence, the unimodal gait feature identification method is prone to environmental disturbance.

On the basis of the theory of pattern classification, we propose an identity-recognition scheme based on adaptive recognition of a person's gait features, as shown in Fig. 6. This method explores the strong generalization ability of K-Nearest Neighbor (KNN) as a classifier and the ability of CNN to fully learn the features of objectives, and it adaptively classifies and learns the gait information of different types of gait features in preprocessed data. The scheme has three steps. Different types of gait features in the preprocessed gait data are classified by KNN, and the results are used as the input data to train the CNN-GR model to extract the feature data in different gait sequences. A softmax classifier is used to classify the extracted gait feature data, a person's identity is determined based on the classification results, and the final results are output.

KNN-based Gait Pattern Classification (KNN-GPC) algorithm is described by Algorithm 2, whose time complexity is $O(k \cdot n)$, where k is the dimensionality of the sample and n is the number of datasets.

5.2.2 Identity-recognition method based on multimodal gait feature fusion

We combine the unimodal gait identification algorithm based on CNN-GR in Section 5.1 and the gait sequence type classification algorithm based on KNN in Section 5.2.1 to realize an adaptive Multi-feature Gait Information based Identification (MGII), described by Algorithm 3.

The time complexity of MGII mainly comes from the time to train the CNN-GR model, and can be approximated by $O\left(\sum_{l=1}^{D} M_l^2 \cdot K_l^2 \cdot C_{l-1} \cdot C_l\right)$.

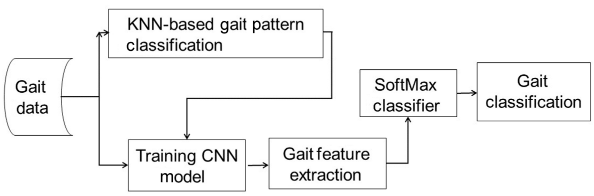


Fig. 6 Gait feature fusion-based identity-recognition scheme.

Algorithm 2 KNN-GPC

Require: Time series of gait data $G(K)$
Ensure: Gait type results
Flag_Classfy = Flase;
if Flage_Classfy **then**
Flag_Classfy = True;
$G'_{x1} = \{G'_{x1}(k) \mid k = 1, \dots, K\}$; //Segment gait sequence x-, y-, z- axis to be measured
Sim $\left(G'_{x1}, G'_{x2}\right)$; //Calculate the similarity of gait sequence samples
T_G = TypeOf $\left(G'_{x2}\right)$; //Mark gait type
end if
return Type_Result //Gait type results

6 Experimental Analysis

6.1 Experimental data

We used two synthetic datasets in an experiment to validate the performance of the CNN-based gait recognition model and the proposed identity algorithms. The first dataset, MIT-Gait, combines the MIT reality mining dataset[19] and our collected crowd gait information. The second dataset, Infocom06-Gait, combines the Infocom2006 trace dataset[20] and our collected crowd gait information. The MIT dataset collects traces and identities of 97 MIT students and faculty. The Infocom06 dataset collects trace and identity information of participants of the Infocom conference. The synthetic MIT-Gait and Infocom06-Gait datasets yielded by data fusion are multimodal.

The Actitracker gait dataset published by WISDM is used in the training stage of CNN-GR[21]. The ONE simulation platform, which is widely used on the mobile internet, is used as the experimental simulation platform. The recognition accuracy, confusion matrix, and kappa statistic[22] are used for performance evaluation.

A Back Propagation Neural Network (BPNN)[23,24] adjusts the network parameters through input-output pairs. It is highly flexible, inductive, and capable of nonlinear modeling, and thus is suitable for processing personal gait information. LBNet[25] is a method that uses CNNs to calculate GEI similarity. It uses the convolutional layer and the additional layer to simulate the weighted subtraction of GEI, and then uses a

Algorithm 3 MGII

Require: Gait training set X and time series of gait data $G(K)$
Ensure: Identification result
CNN_Model = SGFI(X); //Call Algorithm 1 to train the CNN model
Type_Result = KNN-GPC($G(K)$); //Call Algorithm 2 to get gait type results
Type_CNN = CNN_Model(Type_Result); //Use CNN model to recognize gait
return Get_Out //Identification result

multilayer neural network to extract gait features. In the experiment, the MGII and SGFI proposed in this paper are compared with the methods based on BPNN and LBNet, respectively, in terms of the ecognition accuracy, confusion matrix, and kappa statistic.

6.2 Experimental results and analysis

Figure 7 compares the recognition accuracies of the two synthetic fused datasets and the number of trainings. The experimental results show that for the two synthetic datasets, the accuracies of the proposed identity-recognition methods are always above 70%. As the number of CNN training iterations increases, the recognition accuracies for all the datasets increase accordingly. At 10 iterations, for the two multimodal feature datasets, the recognition accuracies of the proposed methods can exceed 80%. These experimental results show that MGII, a CNN-based gait recognition model, achieves relatively good recognition accuracy for data yielded by fusing the gait information of different persons and the multimodal information of persons with different behaviors, and it has a relatively strong recognition capability.

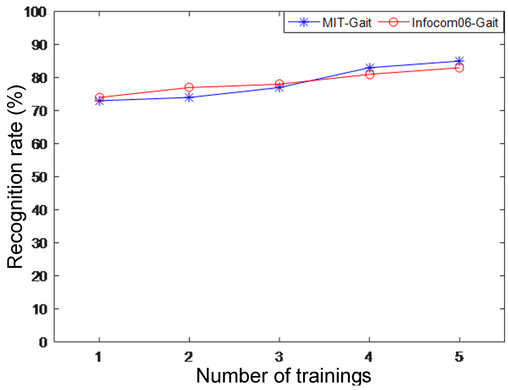


Fig. 7 Recognition rate varies with the number of trainings.

Figure 8 compares the recognition accuracy of MGII, CNN-based, and the BPNN-based gait identification methods. The deep learning based MGII method shows a significant improvement in recognition accuracy over the other methods. The BPNN-based gait recognition method performs worst, with the lowest average recognition scores for the two datasets. Because of the advantages of the CNN, MGII outperforms BPNN and CNN. Because the MGII identity-recognition method first classifies the gait type and then relies on multimodal features to determine a person's identity, its identification accuracy is better than that of CNN.

The confusion matrix and its computation scheme[26, 27] are used to visually compare the performance of the proposed algorithms to that of a supervised learning identification algorithm. Figure 9 shows that the proposed identification algorithms have the better discriminating ability for some actions, indicating that they have a certain practicability. Figures 9a–9d show the confusion matrices of the four algorithms on the MIT-Gait dataset, and Figs. 9e–9h show the results on the Infocom06-Gait dataset. MGII discriminates various gait actions better on both fusion datasets, and thus performs better in-person

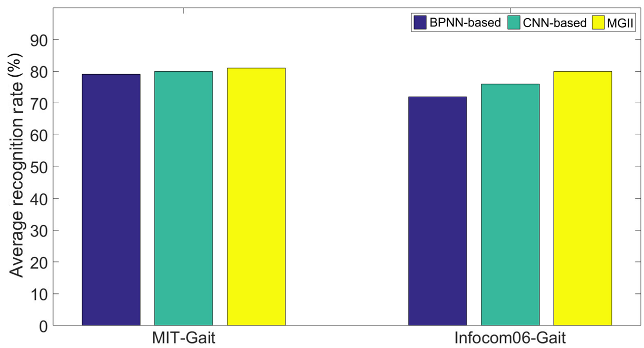
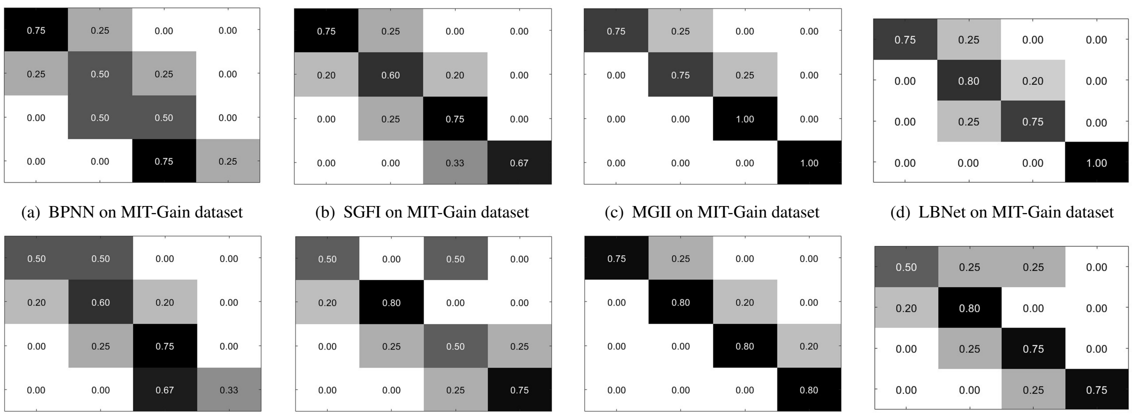


Fig. 8 Comparison of average recognition rates.



(a) BPNN on MIT-Gain dataset (b) SGFI on MIT-Gain dataset (c) MGII on MIT-Gain dataset (d) LBNet on MIT-Gain dataset

(e) BPNN on Infocom06-Gait dataset (f) SGFI on Infocom06-Gait dataset (g) MGII on Infocom06-Gait dataset (h) LBNet on Infocom06-Gait dataset

Fig. 9 Confusion matrix comparison of different algorithms.

identification.

On the basis of the comparison of the confusion matrices, we summarize the kappa statistic for each algorithm. Generally, the kappa statistic is used to evaluate the difference between the classification results of a classification model and the result by random classification. The kappa value locates in the range of $[-1, 1]$. The model classification result is completely different from that of random classification when kappa is 1; in contrast, the classification model has no effect when kappa is 0, resulting in a model classification result completely different from the random classification result. When kappa is -1, the classification model is completely inferior to random classification. Hence the greater the kappa value is, the better a classifier performs. Table 1, a comparison of the experimental results pertaining to the kappa statistic, shows that for the two multimodal synthetic fusion datasets, the kappa statistic of MGII is closer to 1, meaning that the CNN-GR model and the CNN-GR-based MGII identification algorithm perform well in gait classification and identity-recognition.

7 Conclusion

On the basis of the feature mutual exclusion of gait feature data, we propose a gait recognition scheme and models based on the combination of mobile internet and a CNN. Through multimodal information fusion and the application of a clustering method to classify gait sequences in different contexts, we propose person respective identification algorithms based on unimodal gait features and their fusion. Experimental results show that our CNN-based multimodal gait adaptive identification method has higher recognition accuracy than the other gait identification methods considered herein, and the information fusion based multi-feature identification method is more robust than a single-feature identification method.

Table 1 Kappa indicator comparison.

Dataset	Algorithm	Kappa indicator
MIT-Gait	BPNN	0.792
	SGFI	0.811
	LBNet	0.823
	MGII	0.834
Infocom06-Gait	BPNN	0.781
	SGFI	0.805
	LBNet	0.813
	MGII	0.828

However, several limitations need to be further studied in our future works. The CNN-GR model is relatively complex, and the training time is too long, resulting in difficulty in deployment on computing power-lowering mobile smart terminals, which affects its practicability; multimodal data based identification directs the improvement of noncontact identity technology. The efficient data collection that is helpful and conducive for multimodal identification remains difficult.

Acknowledgment

This work was supported by the Smart Manufacturing New Model Application Project Ministry of Industry and Information Technology (No. ZH-XZ-18004), Future Research Projects Funds for Science and Technology Department of Jiangsu Province (No. BY2013015-23), the Fundamental Research Funds for the Ministry of Education (No. JUSRP211A 41), the Fundamental Research Funds for the Central Universities (No. JUSRP42003), and the 111 Project (No. B2018).

References

[1] M. Haghighat, M. Abdel-Mottaleb, and W. Alhalabi, Discriminant correlation analysis: Real-time feature level fusion for multimodal biometric recognition, *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 9, pp. 1984–1996, 2016.

[2] M. P. Murray, A. B. Drought, and R. C. Kory, Walking patterns of normal men, *J. Bone Joint Surg. Am.*, vol. 46, pp. 335–360, 1964.

[3] G. Johansson, Visual perception of biological motion and a model for its analysis, *Perception & Psychophysics*, vol. 14, no. 2, pp. 201–211, 1973.

[4] Y. LeCun and Y. Bengio, Convolutional networks for images, speech, and time-series, in *The Handbook of Brain Theory and Neural Networks*, M. A. Arbib, ed. Cambridge, MA, UK: The MIT Press, 1995, pp. 276–278.

[5] S. A. Niyogi and E. H. Adelson, Analyzing and recognizing walking figures in XYT, in *Proc. IEEE Conf. Computer Vision & Pattern Recognition*, Seattle, WA, USA, 1994, pp. 469–474.

[6] H. J. Ailisto, M. Lindholm, J. Mantyjarvi, E. Vildjiounaite, and S. M. Makela, Identifying people from gait pattern with accelerometers, in *Proc. Volume 5779, Biometric Technology for Human Identification II*, Orlando, FL, USA, 2005, pp. 7–14.

[7] F. Tafazzoli and R. Safabakhsh, Model-based human gait recognition using leg and arm movements, *Engineering Applications of Artificial Intelligence*, vol. 23, no. 8, pp. 1237–1246, 2010.

[8] J. N. Mogan, C. P. Lee, and A. W. C. Tan, Gait recognition using temporal gradient patterns, in *Proc. 5th Int. Conf. Information and Communication Technology*, Melaka,

Malaysia, 2017, pp. 1–4.

[9] E. S. Bigün, J. Biüun, B. Duc, and S. Fischer, Expert conciliation for multi modal person authentication systems by Bayesian statistics, in *Proc. of Int. Conf. Audio- and Video-Based Biometric Person Authentication*, Crans-Montana, Switzerland, 1997, pp. 291–300.

[10] A. K. Jain, S. Prabhakar, and S. Y. Chen, Combining multiple matchers for a high security fingerprint verification system, *Pattern Recognition Letters*, vol. 20, nos. 11–13, pp. 1371–1379, 1999.

[11] A. K. Jain, S. Prabhakar, and L. Hong, A multichannel approach to fingerprint classification, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 4, pp. 348–359, 1999.

[12] A. Baig, A. Bouridane, F. Kurugollu, and G. Qu, Fingerprint-iris fusion based identification system using a single hamming distance matcher, in *Proc. of 2009 Symp. Bio-Inspired Learning and Intelligent Systems for Security*, Edinburgh, UK, 2009, pp. 9–12.

[13] Y. Li, D. Ming, L. Wang, H. Z. Qi, and B. K. Wan, Gait and face profile fusion based human identification at a distance in video, (in Chinese), *Chinese Journal of Scientific Instrument*, vol. 32, no. 2, pp. 264–270, 2011.

[14] A. I. Bazin and M. S. Nixon, Probabilistic combination of static and dynamic gait features for verification, in *Proc. Volume 5779, Biometric Technology for Human Identification II*, Orlando, FL, USA, 2005, pp. 23–30.

[15] D. K. Wagg and M. S. Nixon, Automated markerless extraction of walking people using deformable contour models, *Computer Animation and Virtual Worlds*, vol. 15, nos. 3&4, pp. 399–406, 2004.

[16] J. Mantyjarvi, M. Lindholm, E. Vildjiounaite, S. M. Makela, and H. A. Ailisto, Identifying users of portable devices from gait pattern with accelerometers, in *Proc. of IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Philadelphia, PA, USA, 2005, pp. ii/973–ii/976.

[17] L. Bruton and P. Ramamoorthy, Maximally flat low-pass transfer function synthesis using continuous and discrete filters, *IEEE Transactions on Circuits & Systems*, vol. 23, no. 8, pp. 510–513, 1976.

[18] N. Stojanović, I. Krstić, N. Stamenković, and G. Perenić, Butterworth transfer function with the equalised group delay response in the maximally flat sense, *Electronics Letters*, vol. 54, no. 25, pp. 1436–1438, 2018.

[19] N. Eagle, A. Pentland, and D. Lazer, Inferring friendship network structure by using mobile phone data, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 36, pp. 15274–15278, 2009.

[20] J. Scott, R. Gass, J. Crowcroft, P. Hui, C. Diot, and A. Chaintreau, CRAWDAD dataset cambridge/haggle (v. 20090529), https://crawdad.org/cambridge/haggle/20090529, 2021.

[21] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, Activity recognition using cell phone accelerometers, *ACM SIGKDD Explorations Newsletter*, vol. 12, no. 2, pp. 74–82, 2011.

[22] J. Cohen, A coefficient of agreement for nominal scales, *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 37–46, 1960.

[23] L. L. Liu, Y. L. Yin, W. Qin, and Y. Li, Gait recognition based on outermost contour, *International Journal of Computational Intelligence Systems*, vol. 4, no. 5, pp. 1090–1099, 2011.

[24] R. Hecht-Nielsen, Theory of the backpropagation neural network, in *Neural Networks for Perception*. San Diego, CA, USA: Academic Press, 1992, pp. 65–93.

[25] Z. F. Wu, Y. Z. Huang, L. Wang, X. G. Wang, and T. N. Tan, A comprehensive study on cross-view gait based human identification with deep CNNs, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 2, pp. 209–226, 2017.

[26] P. Ebert, An extension of rate-distortion theory to confusion matrices, *IEEE Transactions on Information Theory*, vol. 14, no. 1, pp. 6–11, 1968.

[27] D. Simon and D. L. Simon, Analytic confusion matrix bounds for fault detection and isolation using a sum-of-squared-residuals approach, *IEEE Transactions on Reliability*, vol. 59, no. 2, pp. 287–296, 2010.

Changjie Wang received the BEng degree from Changsha University of Science & Technology, China in 2018. He is currently a master student at Jiangnan University, China. His research interests include the Internet of Things (IoTs) and cyber security.

Zhihua Li received the BEng degree from Wuxi Light Industry University, China in 1992, and the MEng and PhD degrees from Jiangnan University, China in 2002 and 2009, respectively. He is currently an associate professor at Jiangnan University. His research interests include network technology, parallel/distributed computing, information security, edge computing, and mobile computing.

Benjamin Sarpong received the BEng degree from Kwame Nkrumah University of Science and Technology, Ghana in 2014, and the MEng degree from Jiangnan University, China in 2020. He is currently an experimentalist at Jiangnan University. His research interests include sensor network, mobile computing, and their information security.