Human Identification Based on Gait Manifold

Xiuhui Wang[1*] and Wei Qi Yan[2]

[1*]Key Laboratory of Electromagnetic Wave Information Technology and Metrology of Zhejiang Province, China Jiliang University, Hangzhou, 310018, Zhejiang, China.
[2*] Auckland University of Technology, Auckland, 1010, New Zealand.

*Corresponding author(s). E-mail(s): wangxiuhui@cjlu.edu.cn; Contributing authors: weiqi.yan@aut.ac.nz;

Abstract

Gait-based pedestrian identification has important applications in intelligent surveillance. From anatomical viewpoint, the physical uniqueness of human gait is physiological discriminative of individuals. Therefore, in theory, like fingerprint and face, gait is used as a biometric for pedestrian identification. However, gait-based pedestrian identification still faces multiple challenges due to a vast diversity of walking conditions and complex acquisition environment during data collection. In this paper, through combining the nonlinear dimensionality reduction by using gait manifold and the temporal feature of gated recurrent unit (GRU) together, we propose a novel gait-based pedestrian identification framework. Firstly, we design a temporal enhancement module to construct a series of frame-by-frame gait trend energy images (ff-GTEIs), which represents spatiotemporal gait characteristics and does not reduce the number of samples. Secondly, a supervised locally linear embedding (LLE) dimensionality reduction scheme is proposed, which generates a low-dimensional gait manifold for each pedestrian and transforms all ff-GTEIs into corresponding gait manifold space. Thirdly, a new pedestrian identification network based on residual GRU is proposed, which is able to identify a person by comprehensively considering the similarity between its gait and corresponding gait manifold. Finally, a series of comparative experiments are carried out based on well-known gait datasets, the experimental results show that the proposed framework for pedestrian recognition in this paper exceeds most existing methods, and has achieved an average correct recognition rate 97.4% and 99.6% based on the open-accessed gait dataset CASIA Dataset B and OU-ISIR LP dataset, respectively.

Keywords: Gait recognition, Pedestrian identification, Video surveillance, Gated recurrent unit

1 Introduction

Gait-based pedestrian recognition refers to person identification by detecting physical characteristics and movement posture of pedestrians. As one of the most promising biometric features, compared to fingerprints and facial features, human gait has the advantages of being easy to be collected from long distances, non-offensiveness, etc. [1–6]. Due to influence of pedestrians and complex external factors in data acquisition process, there are many challenges in gait recognition. The most influential factor is the changes of viewing angle and pedestrian dress. In the last decade, a lot of related work was proposed for resolving the problems [7, 8]. However, in the era of big data, the image data collected in many applications is high-dimensional. The increase of data dimensionality facilitates pattern classification and human identification, because more information is available. Nevertheless, when the dimensionality is increased to an extent, a so-called curse of dimensionality will occur. The main reason is that as data dimensionality increases, the data for model training will become sparse, correspondingly, the hyperplane for pattern classification will lead to overfitting or underfitting. In addition, if the data dimensionality is too high, considering the number of samples and the computing power of the computer, it is usually necessary to reduce the dimensionality of data in advance. Principal component analysis (PCA) [9] is currently the most popular linear dimensionality reduction method, but for nonlinear dimensionality reduction, it is beyond its ability for dimensionality reduction. For a high-dimensional gait dataset, it is more suitable to use nonlinear dimensionality reduction methods [10] like manifold learning.

In this paper, we propose a nonlinear dimensionality reduction method, called Gait Manifold for high-dimensional gait datasets. The aim of manifold learning for dimensionality reduction is to find a mapping from high-dimensional manifold to a low-dimensional space so as to represent the data points in high-dimensional space and achieve the purpose of data dimensionality reduction. Typical manifold learning algorithms are ISOMAP [11], LLE [12], Laplacian Eigenmaps [13], etc. In addition, by introducing metric to learn the inter-sample distance in each gait manifold space, we propose a new gait-based pedestrian recognition algorithm. The contributions of this work are fourfold, which are described as follow:

- A novel gait representation, frame-by-frame gait trend energy images (ff-GTEIs) is proposed in this paper. We design a temporal enhancement module to construct a series of frame-by-frame gait trend energy images (ff-GTEIs), which represent spatiotemporal gait characteristics and keep the number of samples at the same time.

- A new generative gait model, called gait manifold is put forward. Gait manifold is a low-dimensional nonlinear manifold learned by considering various factors, such as view angles and clothing changing. By utilizing nonlinear dimensionality reduction, we construct a gait manifold for each person in the given gait dataset.
- Gait-based pedestrian identification network is proffered. A new pedestrian identification network based on residual gated recurrent unit is proposed and trained, which identifies a person by comprehensively considering the similarity between its gait and each gait manifold.
- An advancement in terms of correct recognition rates based on two open-accessed gait datasets is brought up. With the CASIA gait Dataset B and OU-ISIR large population dataset, a series of comparative experiments are conducted, which demonstrate that the proposed methods reach the state-of-the-art requirements.

The rest of this paper is organized as follows. In Section 2, a comprehensive review of the existing gait recognition methods will be presented. Then, in Section 3, we will describe how to achieve nonlinear dimensionality reduction through gait manifold and propose a novel gait-based pedestrian identification algorithm. Next, a series of experiments will be conducted with two open-accessed gait datasets in Section 4. Finally, the conclusion of this paper will be drawn in Section 5.

2 Related work

2.1 Dimensionality Reduction

The existing research work on dimensionality reduction and subspace is grouped into linear dimensionality reduction and nonlinear dimensionality reduction. The former is represented by using principal component analysis (PCA), meanwhile the latter is most popular by using LLE [12, 14], ISOMAP [11, 15] and Laplacian Eigenmaps [13, 16].

As a widely used linear dimension reduction method, PCA and its variants [9] are simple to implement and efficient to discover the intrinsic features on linear subspaces from the high-dimensional input samples. Xiao and Zhou [17] proposed a quaternion ridge regression model for two-dimensional Quaternion Principle Component Analysis (QPCA), which was insensitive to outliers and the robustness of classification. The limitation of this method is that because it works in the column direction of color images, only the bottom row pattern can be found after sparse projection. To minimize the reconstruction error and realize rotational invariance, Wang et al. [18] presented a generalized robust metric learning for PCA, which employed $\ell_{2,p}$ -norm as the distance metric for reconstruction error. Jiang et al. [19] proposed a superpixelwise PCA approach to learn from the intrinsic low-dimensional features of hyperspectral images (HSIs), which carried out dimensionality reduction through a unified projection for an entire HSI and found the low-dimensional linear feature of HSIs. However, due to the complex collecting conditions, gait features do not satisfy this constraint as they generally lie on nonlinear manifolds in the high-dimensional image space. In our previous research work, we applied the improved PCA algorithm [20] to the preprocessing process of gait recognition, but the improvement is limited. Fortunately, manifold algorithms have emerged in recent years, which are based on eigen decomposition and integrate the main characteristics of PCA and other linear algorithms to a class of extensive nonlinear manifolds. These algorithms usually have better computational efficiency, global optimality, and stable convergence.

As a typical representative of nonlinear dimension reduction methods, LLE [12] can learn locally linear low-dimensional manifolds of any dimension, and its computational complexity is relatively small. The main disadvantage of this method is that the target manifold can only be unclosed and different nearest neighbor parameters have a great impact on the final dimension reduction results.

In order to find meaningful low-dimensional features in high-dimensional observations, Tenenbaum et al. [11] proposed a global geometric framework of nonlinear dimension reduction to calculate the global optimal solution and ensure that it converges to the real structure of the original high-dimensional data.The limitation of this method is that it can't get a conclusion directly in the process of calculation, and the final result can only be obtained through multiple matrix transformations, which leads to its low efficiency. Belkin and Nyogi [13] proposed an Laplacian eigenmaps method, which built weighted adjacency matrix with Gaussian kernel and computed embedding from normalized Laplacian. The main disadvantage of this method is that it assumes each point lies in the convex hull of its neighbors and thus it might have trouble at the boundary.

2.2 Gait Classification and Recognition

In this section, we split the existing gait classification and recognition methods into generative methods and discriminative methods, and thus make comprehensive analysis and evaluation [8]. From the perspective of probability distribution, for a group of samples, each of which has a feature and a corresponding class label, the discriminative model is trained by using the conditional probability directly, meanwhile, the generative model learns from the joint probability distribution $P(x, y)$ and is calculated by using the conditional probability.

The generative methods for gait recognition is to train a deep neural network model for each person by using a given database, the corresponding model of all classes are applied to determine the matching probability of query samples [21–26]. In [21], a hidden Markov model was employed to analyse gait motions, the posterior probabilities from the HMM are employed to infer the gait. Chen et al. [22] proposed a framework based on factorial hidden Markov model (FHMM) to solve the problem of feature fusion for gait classification, which has a multilayer structure and provides an alternative way to combine gait features together without concatenating them to a single feature. Uriel et al. [23] propounded a robust probabilistic framework for predicting gait events from human walking, which combines the outputs from a Bayesian network over time to predict the most probable gait event.

Furthermore, by combining optical flow of gait history images, Chen et al. [24] presented a gait representation method, which implemented the localized representation of human motion, and proposed a HMM-based gait recognition method by using two groups of hidden Markov models to distinguish similar gait sequences. Yuan et al. [25] proposed a gait recognition method based on Fourier descriptors and canonical time warping, which firstly utilizes Fourier descriptors to extract gait silhouette feature and then applies canonical time warping to recognize the gait sequence. Deepak et al. [26] analysed gait sequences using latent Dirichlet allocation to recognize human actions, which transforms the extracted gait sequences in gait domain to text domain, and groups the input documents into finer clusters approximately 8-9 clusters.

In short, generative model-based methods represent the distribution of data from a statistical point of view and reflect the similarity of the data itself. The main advantage of the generative methods for gait recognition is that the methods converge very fast, that is, if the sample size increases, the models will converge to the real model quickly, and if there is a hidden variable, the generative methods can still be used. The shortcoming of these methods is that the training process is much complicated.

The discriminative methods for gait recognition were used to train the models from historical walking data, the trained models were applied to calculate the probability that the given sample belongs to a class [27–32]. Nithyakani et al. [27] proposed a scheme for gait recognition to extract the gait features of a person through gait energy image. Alireza [28] proposed a view-invariant scheme for gait recognition which firstly extracts gait convolutional energy maps (GCEM) from frame-level convolutional features, then adopts a bidirectional recurrent neural network to exploit the relationships between spatiotemporal representations, and finally used an attention model to select important partial representations as the information. Upadhyay et al. [29] proposed a biometric method by using gait analysis from visual features which is used for contactless biometric authentication.

Besides, Zou et al. [30] put forward a gait recognition method based on deep neural networks, which successively abstracts gait features in the spatiotemporal domains by using CNN and RNN. Chen et al. [31] utilized a multi-view gait generative adversarial network (MvGGAN) to generate fake samples and extended the existing gait datasets so as to provide adequate gait samples for cross-view gait recognition. The proposed method was used to train a single generator for all view pairs involved in single or multiple datasets, and fulfill

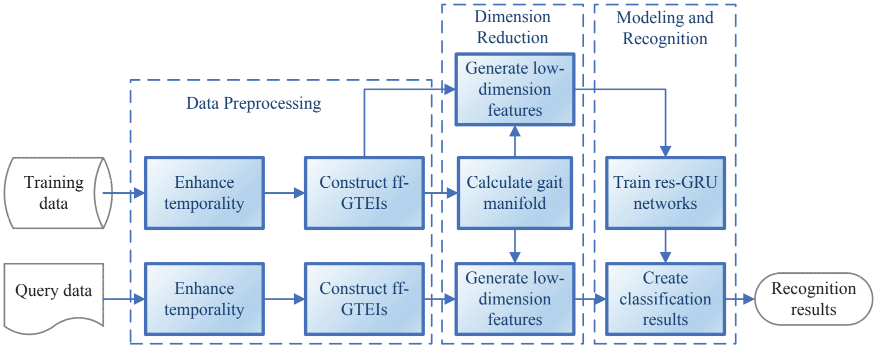

Fig. 1 Flow chart of the proposed human identification framework

domain alignment based on projected maximum mean discrepancy to reduce the influence of distribution divergence caused by sample generation. Hagui et al. [32] presented a new discriminative method for gait recognition by using hybrid conditional random fields (CRF), which used a hidden CRF model to combine two classifiers: A spatial classifier which assigns a label to a local feature and temporal classifier which takes use of a motion history image. The proposed framework firstly extracted spatiotemporal cues from human silhouette, and applies MLP to the two sets of features to obtain the hidden CRF input for gait recognition.

In short, discriminative gait recognition methods are to establish a discriminant function with finite samples, and to find the optimal classification between different classes. The outstanding advantage of this type of methods is that it can clearly distinguish the differences between multiple classes. The main disadvantage is that it does not reflect the characteristics of the training data itself.

The proposed method in this paper belongs to the second category. Firstly, in gait data representation, in order to obtain the intrinsic characteristics with high resolution, we improve conventional gait energy image (GEI) representation [33], and design a new enhancement module to construct a series of ff-GTEIs, which better reflect the spatiotemporal characteristics of gait without reducing the number of samples. Then, we create a manifold-based gait model for each person by using a large number of training samples with various interference factors, we propose a residual GRU network for pedestrian identification. To the best of our knowledge, our proposed gait manifold has not been intensively studied in human gait identification before.

3 Our Method

As shown in Figure 1, the proposed human identification framework consists of three stages:

Fig. 2 The examples of normalized gait silhouette images

1. Data preprocessing. During this stage, the training data and test data are processed through an enhancement module, and corresponding ff-GTEIs are constructed, which is able to represent spatiotemporal gait characteristics.
2. Dimensionality reduction. A supervised local linear embedding dimensionality reduction scheme is proposed, which generates a low-dimensional gait manifold for each pedestrian and transforms all ff-GTEIs into corresponding gait manifold space.
3. Modelling and recognition. A new pedestrian identification network based on residual GRU is proposed and trained, which identifies a person by comprehensively considering the similarity between its gait and the gait manifold.

In the rest of this paper, we will discuss on details the key method of this paper, i.e., gait representation, gait manifold, residual GRU, network modelling, and gait recognition.

3.1 Gait Representation

The data preprocessing stage includes two steps: The enhancement of temporal characteristics in gait silhouette sequences and the ff-GTEI generating. The input data in this paper is normalized gait silhouette images as shown in Figure 2. Based on this, firstly, gait silhouette images are transformed into instantaneous trend image (tagged as ITI) sequence through a temporal difference network, and then ff-GTET images are established by computing arithmetic mean of adjacent trend images.

As shown in Figure 3, ITIs are generated by calculating the difference between adjacent frames of gait silhouette videos, which reflect the instantaneous movement of human body in the process of walking.

After obtained the ITI sequences of each person, we construct the corresponding ff-GTEIs images. Given an ITI sequence $\{I_1, I_2, \cdots, I_K\}$, the ff-GTEI corresponding to the t-th frame is defined as

$$G_t = \frac{1}{P} \sum_{i=t}^{t+p} I_i\left(u, v\right), \tag{1}$$

where P is the span value which determines the time span of the dynamic gait features captured by using ff-GTEIs, I_i is the i-th ITI, t is the frame number in current ITI sequence, and (u, v) is the 2D image coordinate.

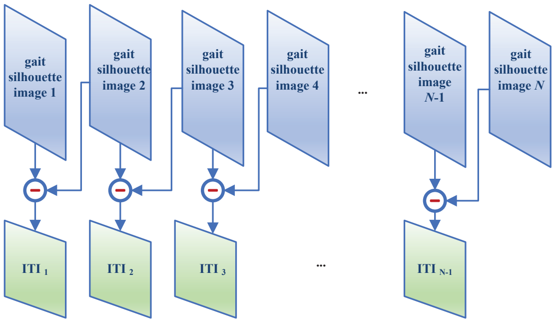

Fig. 3 The construction of instantaneous trend images

Compared with gait silhouette images and gait energy images in gait recognition, ff-GTEI has the following advantages:

- It reflects the dynamic characteristics of cross natural gait cycle. The traditional gait energy image is based on natural gait cycles and can only reflect the dynamic gait characteristics in each gait cycle;
- It keeps the number of samples almost unchanged. On the contrary, traditional gait energy image is generated by merging the gait contour images in a gait cycle, which will greatly reduce the number of samples;
- The problem of using sequential input or random sequence input of training samples is solved. While using gait silhouette images as training data, the model training will be affected if the sequence input is used, but the integrity of dynamic gait features cannot be maintained. The ff-GTEI not only ensures the training outcomes by random input in model training, but also keeps the integrity of dynamic gait features.

3.2 Gait Manifold

Given a set of ff-GTEIs of the p-th person under the c-th condition as ϑ_p^c, which is defined as:

$$\vartheta_p^c = \{I(p,c) \mid c \in C\}, \tag{2}$$

where $I(p,c)$ is an ff-GTEI of the p-th person under the c-th condition, c denotes one of such factors such as view angles, clothing, carrying a bag, etc. Considering the ff-GTEIs of the p-th person, we obtain the gait manifold ϕ_p corresponding to the p-th person as

$$\phi_p = \bigcup_c \vartheta_p^c \tag{3}$$

where C refers to a combination of influencing factors during the process of gait sample acquisition.

As shown in Figure 4, considered a group of gait silhouette images collected by using four cameras from various view angles during individual walking,

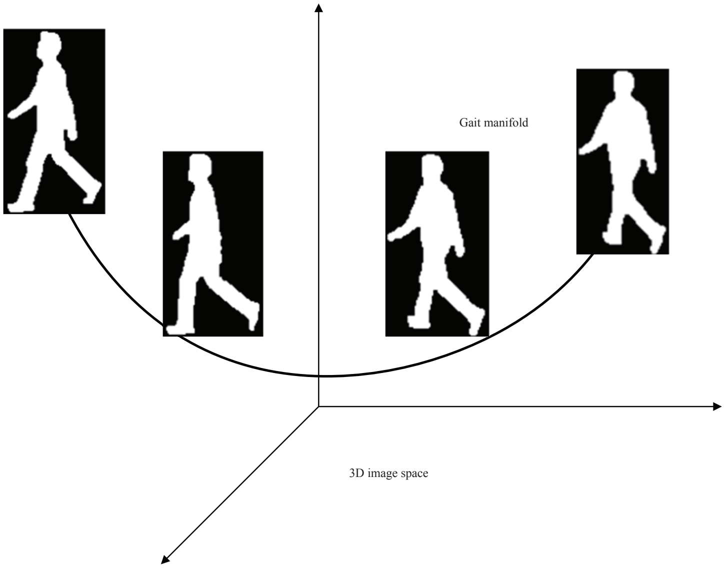

Fig. 4 Schematic diagram of cross-view gait manifold embedding

the only degree of freedom in this case is the rotation angle of walking direction. Therefore, the intrinsic feature dimensionality of such gaits is very small. However, because the size of these silhouette images is $150 \times 80 = 12{,}000$ pixels, these intrinsic gait features are embedded into a 12,000-dimensional image space.

Combining with the actual scene of gait-based human identification, we propose a supervised LLE method to generate each person's gait manifold. Using gait manifold for supervised dimensionality reduction of ff-GTEI includes two steps:

1. For the training data, only the corresponding gait manifold is used for dimensionality reduction, and the reduced data is applied to learn and discover the hidden low-dimensional nonlinear gait manifold of each individual. Considering the time cost of dimensionality reduction, we randomly select k samples from each person's ff-GTEIs to generate the corresponding gait manifold, and then transform the rest of the samples into the corresponding gait manifold space. Among them, k as a super parameter will be optimized in the training process.
2. For the query data, we transform the unknown samples into each known gait manifold space in turn, and obtain N low dimensional gait feature vectors (N is the total number of types) for gait classification in the later stage.

3.3 Residual GRU and Gait Recognition

Long short-term memory (LSTM) and GRU are two mainstream RNNs, which can learn long-term dependencies through memory gates based on time series

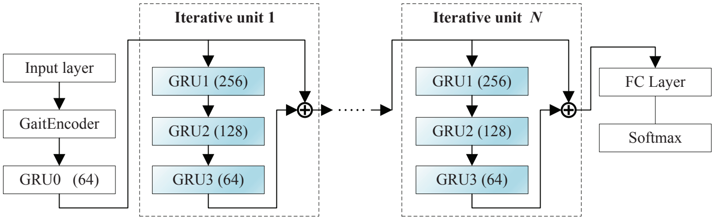

Fig. 5 The proposed residual gated recurrent unit network model

and thus solve the problems of vanishing gradient and exploding gradient in the RNN models.

In addition, compared with LSTM, GRU has fewer parameters and faster convergence speed, so it can greatly accelerate our iterative process. Therefore, we use deep GRU network for gait modelling. Furthermore, in order to solve the degradation problem of deep GRU network, that is, when the depth of GRU network is large, with the further increase of the number of model layers, the correct recognition rate may decrease, we add residual branches into the proposed GRU-based gait model. Compared with the traditional GRU networks, the res-GRU model proposed in this paper adopts three GRU layers plus a shortcut connection to form a basic unit, and finally takes use of N basic units to form the backbone of the whole model. In addition, after the input layer, res-GRU adds a GaitEncoder layer to preprocess the input ff-GTEI data.

As shown in Figure 5, the proposed residual GRU network consists of an input layer, a GaitEncoder layer, a GRU_0 layer, a fully connected (FC) layer, a softmax layer, and a serial of iterative units. Each iterative unit is comprised of three GRU layers, including 256, 128, and 64 hidden cells, respectively, and a shortcut branch. The GaitEncoder layer is utilised to convert the input low-dimensional ff-GTEIs into triples (batch-size, time-step and input-size) which will be used as input of the GRU0 layer. When gait feature information propagates forward through each iterative unit, the input information is added to the output of the GRU3 Layer via a shortcut connection.

In gait recognition, corresponding to each ff-GEI of the query gait data, we are use of all the known gait manifold to generate a set of low-dimensional gait feature. Then, a series of classification results are obtained by inputting the gait features into the trained res-GRU network. At last, we select the best label corresponding to the class with the maximum probability as the final result.

The details of gait-manifold-based human identification are specified in Algorithm (1) and Algorithm (2), respectively.

Algorithm 1. Gait manifold construction and model training

Input: Training data.
Output: N gait manifolds and a res-GRU gait model.
Step 1: Compute instantaneous trend images of the training data by temporal enhancement.

Step 2: Construct ff-GTEI for all training data.
Step 3: Create gait manifold for each class, and obtain N gait manifolds.
Step 4: Train res-GRU model by Back propagation algorithm [34].

Algorithm 2. Gait-based human identification

Input: Test data, N gait manifolds and a res-GRU gait model.
Output: Identification result.
Step 1: Compute instantaneous trend images of the query data by temporal enhancement.
Step 2: Construct ff-GTEIs from given query data, and obtain M ff-GTEIs.
Step 3: Transform the M ff-GTEIs into each gait manifold in turn and obtain N groups of low-dimensional gait features.
Step 4: Input N groups of low-dimensional gait features into the res-GRU gait model and get N results with the corresponding probabilities.
Step 5: Select the result with the max probability as the final identification result.

---

4 Experiments

In this section, we design and implement five experiments to evaluate the proposed framework. Firstly, two experiments are carried out on two public gait datasets. By comparing with the existing gait recognition methods, the effectiveness of the proposed recognition framework is verified. Then, we utilize three ablation experiments to demonstrate the contributions of the ff-GTEI representation, gait manifold and res-GRU model, respectively. The experimental environment for this work is Anaconda 1.7.2 plus Pytorch 1.2.0. The experimental platform is Intel® Xeon® CPU E5-2630 V4 (10 cores, 2.4GHz), with two NIVIDA® GeForce® RTX 2080 Ti model graphics cards (11GB video memory) and Centos® 7-9.2009 operating system.

4.1 Gait Datasets

In our experiments, we used two public gait datasets, CASIA Dataset B [35], and OU-ISIR LP Dataset [36], which are widely used in many documents.

The CASIA Dataset B is a large multi-view gait dataset provided by the Institute of Automation, Chinese Academy of Sciences (CASIA). There are 124 subjects in CASIA Dataset B, whose data are collected from 11 perspectives. Besides, three variations, namely, view angles, clothing and carrying conditions, are separately considered. The examples from CASIA Dataset B are shown in Figure 6, where from the first line to the third line, the samples are separately with conditions of normal, wearing overcoat and carrying a bag, from the first column to the third column, the angles between camera direction and walking direction are respectively of 0 degree, 90 degree and 180 degree.

Fig. 6 Examples from CASIA Dataset B

Fig. 7 The examples from OU-ISIR LP dataset

The OU-ISIR LP dataset was collected by the Institute of Scientific and Industrial Research (ISIR), Osaka University (OU), which consists of person walking on the ground surrounded by 2 cameras with 30 fps, 640 $\times$ 480 resolution. The latest version of this dataset includes 4,016 subjects with ages from 1 to 94 years old. Besides, there are two subsets, A and B. Subset A is related to a set of two sequences (gallery and probe sequences) per subject. Subset B is about a set of sequences per subject. In addition, each of the main subsets is further divided into 5 subsets based on the observation angles: 55 degrees, 65 degrees, 75 degrees, and 85 degrees. The examples from OU-ISIR LP dataset are shown in Figure 7.

4.2 Benchmarks and Evaluations

In order to better evaluate the proposed human identification framework, we chose three state-of-the-art methods [2, 7, 8] and typical methods [20, 33, 37] in the comparative experiments.

- Similarity calculation [2]. After comprehensive study on gait-based human identification, an extensive evaluation is proposed in terms of various tasks with multiple preprocessing approaches. Three CNN architectures are investigated by matching local features. In this paper, the most representative LB architecture is selected as one of the comparison methods.

- Task-centred CNN architectures (TCNN) [7]. In this work, the contrastive loss and the triplet ranking loss are utilized for cross-view gait recognition. Effective input/output architectures of CNN are investigated for person identification.
- Convolutional LSTM [8]. In previous work, we present a new gait representation to increase the amount of available training data. Then, in order to extract the intrinsic temporality of human gait, we design a novel gait recognition model by adding convolutional layers to the traditional LSTM network architecture.
- $(2D)^2$PCA-based gait recognition [20]. In order to verify the superiority of nonlinear dimensionality reduction based on gait manifold, we use $(2D)^2$PCA as one of the comparison benchmarks.
- View transformation model (VTM) for gait recognition [37]. As a typical solution of cross view gait recognition, VTM is constructed with a training set of multiple persons from multiple view directions. In our experiment, we implement it as a baseline and compare the performance with other approaches.
- The original GEI-based gait recognition [33]. In this paper, the novel gait feature representation, Gait Energy Image (GEI), is firstly proposed, which are adopted broadly. In this experiment, we apply it to verify the effectiveness of the proposed gait representation.

In addition, to compare different methods quantitatively, we employ two kinds of evaluation criteria, ROC curve and average correct recognition rate (Average CRR). The former is an evaluation criterion of empirical computations, which tests the recognition ability of a classifier according to the samples with a threshold, the latter is combined with standard deviation to evaluate the stability of each algorithm.

4.3 Experiment I: Comparative Experiments on CASIA Dataset B

In order to verify the recognition performance of our method in multi-view scenarios with complex interference factors, we designed a comparative experiment based on CASIA Dataset B. The experimental results on CASIA Dataset B are shown in Table 1 and Figure 8.

From the average CRR and standard deviation in Table 1, we see that, compared with the traditional methods [20, 33, 37], the proposed method has obvious advantages in average CRR and standard deviation, that is, the average CRR is increased by at least 8.9%, the standard deviation is significantly improved. We also see from Table 1 that compared with the methods [2, 7, 8], our proposed method is closer to the best performance in terms of standard deviation, and the average CRR is at least 1.5% higher.

Table 1 Results of Comparative Experiments Based on CASIA Dataset B

Methods	Average CRR (%)	Standard deviation
$(2D)^2$PCA [20]	87.2	0.45
VTM [37]	88.5	0.41
OriginalGEI [33]	81	0.4
LB [2]	92.8	0.46
TCNN [7]	94.3	0.43
Conv-LSTM [8]	95.9	**0.31**
Our method	**97.4**	0.32

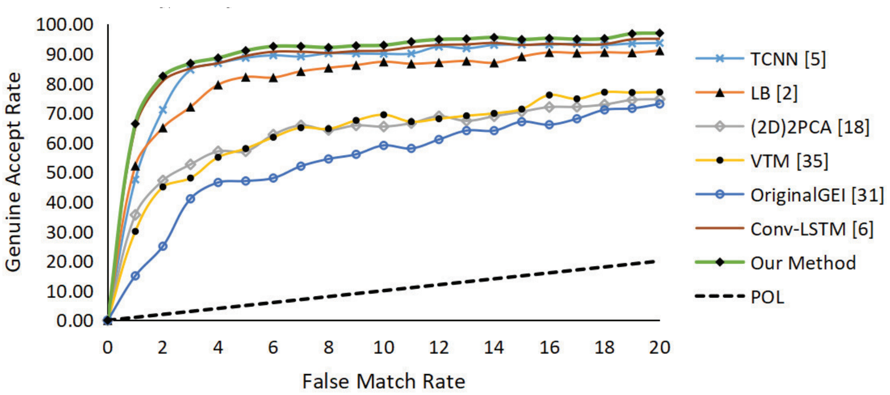

Fig. 8 The ROC curves for comparative experiments based on CASIA Dataset B

Furthermore, from the ROC curves shown in Figure 8, we see that the ROC curve corresponding to the proposed method is close to the upper left corner, which indicates that the recognition ability of this method has reached the most advanced level.

Table 2 The Results of Comparative Experiments Based on OU-ISIR LP Dataset

Methods	Average CRR (%)	Standard Deviation
$(2D)^2$PCA [20]	88.9	0.39
VTM [37]	85.2	0.37
OriginalGEI [33]	84.3	0.29
LB [2]	94.6	0.37
TCNN [7]	98.3	0.34
Conv-LSTM [8]	99.1	**0.23**
Our method	**99.6**	0.25
$(2D)^2$PCA [20]	87.2	0.45
VTM [37]	88.5	0.41
OriginalGEI [33]	81	0.4
LB [2]	92.8	0.46
TCNN [7]	94.3	0.43
Conv-LSTM [8]	95.9	**0.31**
Our method	**97.4**	0.32

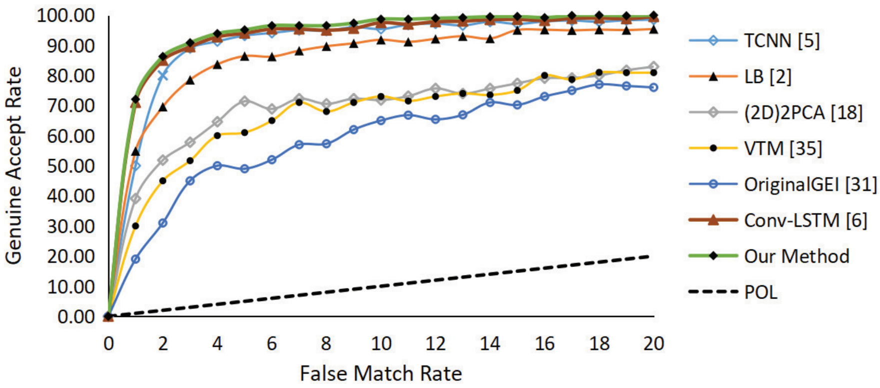

Fig. 9 The ROC curves for comparative experiments on OU-ISIR LP dataset

4.4 Experiment II: Comparative Experiments on OU-ISIR LP

In this section, we designed a comparative experiment based on OU-ISIR LP Datasets to verify the recognition performance of our method on a large-scale dataset with more categories, and the experimental results are shown in Table 2 and Figure 9. From Table 2, we see that the proposed method in this paper has at least 0.5% improvement in average CRR, and is close to the best result in standard deviation. Furthermore, from the ROC curves shown in Figure 9, our method proposed in this paper has better identification outcome.

There are two reasons why our method achieves this performance: 1) We adopt a movement trend-based gait feature representation, which is stable and has stronger ability for feature expression; 2) In this paper, gait manifold is employed for nonlinear dimensionality reduction and optimization of gait features.

4.5 Ablation Experiments

In order to analyse the contribution of gait feature representation, along with our gait manifold and our deep learning model, three ablation experiments were conducted in this paper.

Firstly, in Experiment III, two gait feature representation methods are employed as input while the feature dimensionality reduction and classifier model are unchanged, and the conresponding results are shown in Table 3.

Table 3 Average CRR of Ablation Experiments for Experiment III

Gait representations	CASIA Dataset B	OU-ISIR LP dataset
ff-GEI	97.2	99.5
ff-GTEI	**97.4**	**99.6**

From Table 3, we see that, compared with ff-GEI, which keeps the number of samples almost unchanged, the average CRR is slightly improved if ff-GTEI is used as the input. The main reason is that ff-GTEI not only has the gait features in spatiotemporal domain, but also shows the movement trend, it reflects the gait differences of different persons.

Secondly, in Experiment IV, we take use of the same input and classifier, but the dimensionality reduction module is adopted. The experimental results are shown in Table 4.

Table 4 Average CRR of Ablation Experiments for Experiment IV

Dimensionality reduction	CASIA Dataset B	OU-ISIR LP dataset
PCA	94.1	97.5
$(2D)^2$PCA [20]	95.7	98
Gait Manifold	**97.4**	**99.6**

In Table 4, compared with linear dimensionality reduction methods, the proposed gait manifold can significantly improve average CRR by 2.7% and 1.7% on based CASIA Dataset B, 2.1% and 1.6% on OU-ISIR LP dataset. The reason is that the gait manifold has the ability to discover the hidden low-dimensional nonlinear manifold of individual's gait.

Finally, in Experiment V, we are use of the same input and dimensionality reduction methods, and the experimental results are shown in Table 5.

Table 5 Average CRR of Ablation Experiments for Different Nets

Network architectures	CASIA Dataset B	OU-ISIR LP dataset
LSTM	96.1	99
GRU	96.8	99.2
res-GRU	**97.4**	**99.6**

From Table 5, we see that, the res-GRU recognition model proposed in this paper improves the performance of gait recognition, especially on CASIA Dataset B, which is 1.3% and 0.6% higher than LSTM and pure GRU methods, respectively. The reason is that, by introducing residual module, the input signal is transmitted directly from the input layer to the output layer if it is transmitted forward, so that the degradation problem of deep neural networks is to be resolved. In the case of backpropagation, the errors are directly propagated to the lower layer without any intermediate weight matrix transformation, which alleviates the gradient vanishing problem.

5 Conclusion

In this paper, we proposed a novel gait-based pedestrian identification framework. Firstly, the proposed framework utilities a series of frame-by-frame gait trend energy images to represent spatiotemporal gait characteristics by using a temporal enhancement module. Then, we presented a supervised LLE dimensionality reduction method to generate a low-dimensional gait manifold for each pedestrian and transformed all ff-GTEIs into corresponding gait manifold space. Finally, a new pedestrian identification network based on residual gated recurrent unit is proposed and trained, which identifies a query person by comprehensively considering the similarity between its gait and each gait manifold. The experimental results based on two open-accessed gait datasets show that the proposed framework achieves state-of-the-art performance and unfolds a great potential in identity applications.

6 Acknowledgment

This work was supported in part by Key Research and Development Plan of Zhejiang Province (No.2021C03151) and Natural Science Foundation of Zhejiang Province (No. LY20F020018). The authors would like to thank the Institute of Automation, Chinese Academy of Sciences for providing CASIA gait datasets, and also the Institute of Scientific and Industrial Research, Osaka University for providing the OU-ISIR gait datasets.

References

[1] S. Sarkar, P.J. Phillips, Z. Liu, I.R. Vega, P. Grother, and K.W. Bowyer. The humanID gait challenge problem: Data sets, performance, and analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(2):162–177, 2005.

[2] Zifeng Wu, Yongzhen Huang, Liang Wang, Xiaogang Wang, and Tieniu Tan. A comprehensive study on cross-view gait based human identification with deep CNNs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(2):209–226, 2017.

[3] Hanqing Chao, Kun Wang, Yiwei He, Junping Zhang, and Jianfeng Feng. GaitSet: Cross-view gait recognition through utilizing gait as a deep set. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1, 2021.

[4] Y. Wang, Y. Chen, Mza Bhuiyan, Y. Han, S. Zhao, and J. Li. Gait-based human identification using acoustic sensor and deep neural network. *Future Generation Computer Systems*, 86:1228–1237, 2017.

[5] W. Yan. *Introduction to Intelligent Surveillance.* Springer, 2019.

[6] W. Yan. *Computational Methods for Deep Learning.* Springer, 2021.

[7] Noriko Takemura, Yasushi Makihara, Daigo Muramatsu, Tomio Echigo, and Yasushi Yagi. On input/output architectures for convolutional neural

network-based cross-view gait recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(9):2708–2719, 2019.

[8] Xiuhui Wang and Wei Qi Yan. Human gait recognition based on frame-by-frame gait energy images and convolutional long short-term memory. *International Journal of Neural Systems*, 30(01):1950027, 2020.

[9] Ran He, Bao-Gang Hu, Wei-Shi Zheng, and Xiang-Wei Kong. Robust principal component analysis based on maximum correntropy criterion. *IEEE Transactions on Image Processing*, 20(6):1485–1494, 2011.

[10] Pavan Turaga, Ashok Veeraraghavan, Anuj Srivastava, and Rama Chellappa. Statistical computations on Grassmann and Stiefel manifolds for image and video-based recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(11):2273–2286, 2011.

[11] Joshua Tenenbaum, Vin Silva, and John Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:2319–23, 2001.

[12] Sam Roweis and Lawrence Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science (New York, N.Y.)*, 290:2323–6, 2001.

[13] M Belkin and P Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Journal Neural Computation*, 15:1373–1396, 2003.

[14] Jing Chen and Zhengming Ma. Locally linear embedding: A review. *International Journal of Pattern Recognition and Artificial Intelligence*, 25:985–1008, 2011.

[15] Honggui Li and Maria Trocan. Sparse reconstruction of ISOMAP representations. *Journal of Intelligent and Fuzzy Systems*, 37:1–17, 2019.

[16] Ying Cao and Di-Rong Chen. On the regularized Laplacian eigenmaps. *Journal of Statistical Planning and Inference*, 142:1627–1643, 2012.

[17] Xiaolin Xiao and Yicong Zhou. Two-dimensional quaternion PCA and sparsePCA. *IEEE Transactions on Neural Networks and Learning Systems*, 30(7):2028–2042, 2019.

[18] Qianqian Wang, Quanxue Gao, Xinbo Gao, and Feiping Nie. $\ell_{2,p}$ -norm based PCA for image recognition. *IEEE Transactions on Image Processing*, 27(3):1336–1346, 2018.

[19] Junjun Jiang, Jiayi Ma, Chen Chen, Zhongyuan Wang, Zhihua Cai, and Lizhe Wang. Superpca: A superpixelwise PCA approach for unsupervised feature extraction of hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 56(8):4581–4593, 2018.

[20] Xiuhui Wang, Jun Wang, and Ke Yan. Gait recognition based on Gabor wavelets and $(2D)^2$PCA. *Multimedia Tools and Applications*, 77:1–17, 05 2018.

[21] Joonbum Bae. Gait analysis based on a hidden Markov model. In *International Conference on Control, Automation and Systems*, pages 1025–1029, 2012.

[22] Changhong Chen, Jimin Liang, Heng Zhao, Haihong Hu, and Jie Tian. Factorial HMM and parallel HMM for gait recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 39(1):114–123, 2009.

[23] Uriel Martinez-Hernandez, Mohammed I. Awad, Imran Mahmood, and Abbas A. Dehghani-Sanij. Prediction of gait events in walking activities with a Bayesian perception system. In *International Conference on Rehabilitation Robotics (ICORR)*, pages 13–18, 2017.

[24] Kai Chen, Shiyu Wu, and Zhihua Li. Gait recognition based on GFHI and combined hidden Markov model. In *International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pages 287–292, 2020.

[25] Wei Yuan, Qinkun Xiao, and Liqin Li. Gait recognition based on Fourier descriptors and canonical time warping. In *International Symposium on Computational Intelligence and Design (ISCID)*, volume 1, pages 64–67, 2015.

[26] Deepak N.A., Hariharan R., and U.N. Sinha. Analysing gait sequences using latent Dirichlet allocation for certain human actions. In *National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*, pages 1–4, 2013.

[27] P Nithyakani, A. Shanthini, and Godwin Ponsam. Human gait recognition using deep convolutional neural network. In *International Conference on Computing and Communications Technologies (ICCCT)*, pages 208–211, 2019.

[28] Alireza Sepas-Moghaddam and Ali Etemad. View-invariant gait recognition with attentive recurrent learning of partial representations. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 3(1):124–137, 2021.

[29] Jaychand Upadhyay, Tad Gonsalves, Rohan Paranjpe, Hiralal Purohit, and Rohan Joshi. Biometric identification using gait analysis by deep learning. In *IEEE International Conference for Innovation in Technology (INOCON)*, pages 1–4, 2020.

[30] Qin Zou, Yanling Wang, Qian Wang, Yi Zhao, and Qingquan Li. Deep learning-based gait recognition using smartphones in the wild. *IEEE Transactions on Information Forensics and Security*, 15:3197–3212, 2020.

[31] Xin Chen, Xizhao Luo, Jian Weng, Weiqi Luo, Huiting Li, and Qi Tian. Multi-view gait image generation for cross-view gait recognition. *IEEE Transactions on Image Processing*, 30:3041–3055, 2021.

[32] Mabrouka Hagui and Mohamed Ali Mahjoub. Hidden conditional random fields for gait recognition. In *International Image Processing, Applications and Systems (IPAS)*, pages 1–6, 2016.

[33] J. Han and Bir Bhanu. Individual recognition using gait energy image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(2):316–322, 2006.

[34] J Amrutha and A S Remya Ajai. Performance analysis of backpropagation algorithm of artificial neural networks in verilog. In *IEEE International Conference on Recent Trends in Electronics, Information Communication Technology (RTEICT)*, pages 1547–1550, 2018.

[35] Shiqi Yu, Daoliang Tan, and Tieniu Tan. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In *International Conference on Pattern Recognition (ICPR'06)*, volume 4, pages 441–444, 2006.

[36] Haruyuki Iwama, Mayu Okumura, Yasushi Makihara, and Yasushi Yagi. The OU-ISIR gait database comprising the large population dataset and performance evaluation of gait recognition. *IEEE Transactions on Information Forensics and Security*, 7(5):1511–1521, 2012.

[37] Shuai Zheng, Junge Zhang, Kaiqi Huang, Ran He, and Tieniu Tan. Robust view transformation model for gait recognition. In *International Conference on Image Processing*, pages 2073–2076, 09 2011.