Human Gait Recognition Based on Self-Adaptive Hidden Markov Model

Xiuhui Wang, Shiling Feng, and Wei Qi Yan

Abstract—Human gait recognition has numerous challenges due to view angle changing, human dressing, bag carrying, and pedestrian walking speed, etc. In order to increase gait recognition accuracy under these circumstances, in this paper we propose a method for gait recognition based on a self-adaptive hidden Markov model (SAHMM). First, we present a feature extraction algorithm based on local gait energy image (LGEI) and construct an observation vector set. By using this set, we optimize parameters of the SAHMM-based method for gait recognition. Finally, the proposed method is evaluated extensively based on the CASIA Dataset B for gait recognition under various conditions such as cross view, human dressing, or bag carrying, etc. Furthermore, the generalization ability of this method is verified based on the OU-ISIR Large Population Dataset. Both experimental results show that the proposed method exhibits superior performance in comparison with those existing methods.

Index Terms—Human gait recognition, self-adaptive hidden Markov model, biometrics, video-based surveillance

1 Introduction

Gait recognition is one of the most promising biometric technologies [1], which can be used in intelligent video surveillance and other applications. Gait recognition utilizes physical characteristics and movement posture of a pedestrian [2]. This process involves various complex factors, including motor nerve sensitivity, muscle strength, body structure, and walking habits, etc. Compared with other biometrics, such as facial features, fingerprint and iris, gait has a myriad of advantages such as non-offensive, low-resolution, and easy data collection [3].

In recent years, video-based gait recognition has gained tremendous attention [4], [5], [6], [7], [8] from multiple fields. Nowadays, the existing gait recognition methods mainly focus on variations of view angles, also known as the cross-view problem. However, the existing solutions are far away from completely solving this problem because the gait of one person can be dramatically altered when the view angle changes.

On the other hand, as an outstanding representative of traditional pattern recognition algorithms, HMM models [9] have been successfully applied to solve time series problems, such as natural language processing [10], object tracking [11], etc. Deng, et al. [10] proposed a sparse HMM-based speech enhancement method using a single channel that deals with speech and noise gains accurately in non-stationary and noisy environments. In [11], an adaptive HMM was proposed to carry out video-based face recognition. Although HMM and its variations have been extensively applied to human behavior analysis, natural language processing, object tracking and recognition, etc., few of them was employed to resolve video-based gait recognition. In addition, though currently deep learning techniques are widely used in various image classification and recognition, deep learning-based methods are data driven and usually require a huge amount of training data. However, compared with large data in other applications, the existing gait datasets can only be regarded as small ones. On these small datasets, the advantages of deep learning technologies are not obvious.

Therefore, in this paper, we propose a novel method based on self-adaptive HMM (SAHMM) to promote the accuracy of human gait recognition. First, we construct and initialize a SAHMM gait model for each individual in a given dataset. Then, a new gait representation is constructed so as to train the gait models. In the training process, we propose an incremental and adaptive method to refine the parameters of each SAHMM-based gait models. Finally, gait recognition is achieved by calculating the average probability of all models with respect to an unknown gait. To the best of our knowledge, this is the first time that this approach is proposed. The contributions of this paper include the following three-fold:

1) *A well-designed gait representation local gait energy image (LGEI) and its extraction algorithm.* As an extension of the traditional GEI gait feature representation, LGEI pays much attention to the relationship between adjacent frames and local details of human walking to reduce the influence due to view angle changes.
2) *A new SAHMM-based method for human gait recognition*. By taking advantage of HMM models, we propose a self-adaptive hidden Markov model and the corresponding gait recognition algorithm. In this method, using a small number of samples with similar acquisition conditions to the target environment, an incremental and adaptive method is employed to refine the parameters of each gait model.

- *X. Wang and S. Feng are with the Key Laboratory of Electromagnetic Wave Information Technology and Metrology of Zhejiang Province, Jiliang University, China. E-mail: {wangxiuhui, fengshiling}@cjlu.edu.cn.*
- *W.Q. Yan is with the Auckland University of Technology, Auckland 1010, New Zealand. E-mail: wyan@aut.ac.nz.*

Manuscript received 14 Nov. 2018; revised 27 Sept. 2019; accepted 30 Oct. 2019. Date of publication 4 Nov. 2019; date of current version 3 June 2021. (Corresponding author: Xiuhui Wang.)
Digital Object Identifier no. 10.1109/TCBB.2019.2951146

3) *Advance of recognition rates based on two open-accessed gait databases.* We conduct a series of comparative experiments based on the CASIA database and OU-ISIR large population database, which demonstrate that our method works well with large-scale datasets having cross-view problems.

The remaining sections of this paper are organized as follows. In Section 2, the related work on gait recognition will be introduced. In Section 3, the SAHMM-based gait recognition method will be presented and explained on details including gait feature extraction, parameter estimation and adaptation. Our experiments based on the CASIA gait dataset are conducted to evaluate the proposed algorithms in Section 4; our conclusion and future work of this paper will be finally addressed in Section 5.

2 Related Work

Most of the existing approaches for gait recognition can be classified into two categories, i.e., GEI-based holistic features [3], [5], [12], [13], [14], [15] and details-based features [6], [7], [8], [16], [17], [18], [19]. Gait energy image (GEI) [12] is a spatiotemporal feature representation which is used to characterize human walking gestures. The first kind of methods are based on the GEI and its varieties which generally do not need to extract specific gait features or construct 3D models. In [3], through the GEI, the problem of cross-view gait-based human identification was probed by using deep convolutional neural networks. Connie et al. [5] combined a multi-view matrix representation and a randomized kernel extreme learning machine together, proposed an end-to-end solution for the view-variation problem under Grassmann manifold. In order to reduce the deterioration due to covariation in gait recognition and classification, Guan et al. [13] proposed a class of ensemble methods, which are sensitive to the location of a gait and could not generalize well under most conditions. More recently, Islam et al. [14] proposed a wavelet-based method for feature extraction and designed a gait representation in frequency domain, which was an extension of the GEI. Wang et al. [15] utilized a method based on two-dimensional principal component analysis to reduce the dimension of feature space, a Gabor wavelet-based algorithm for gait recognition was employed to address the problem related to view changing. In short, the gait recognition methods based on GEI-like holistic features utilize the spatiotemporal features of gait silhouette images to differentiate gaits from different persons. The features extracted by using this type of methods generally have higher dimensions, so it is necessary to perform dimensionality reduction in advance. In addition, because this type of methods ignore the details of gait features and cross-cycle characteristics of pedestrian walking, it will affect the accuracy of the subsequent gait recognition. As a summary, gait recognition methods based on detailed features mainly utilize the structural and geometric features of human walking. This type of methods usually require gait videos acquired simultaneously with multiple viewing angles to reconstruct the 3D model. Therefore, its requirements for training data collection are very high, which are difficult to be satisfied by most existing databases and practical applications.

Other methods make use of detailed gait information mainly including structural features and geometric features. Luo et al. [6] utilized a clothes-independent 3D parametric model to deal with the gait variations in speed, inclined plane and clothing. In [7], a variation of Gait Energy Images, i.e., frame-by-frame GEI (ff-GEI), was proposed to expand the volume of available GEI data and relax the constraints of gait cycle segmentation required by most gait recognition methods. Tang et al. [8] assumed that a 3D object shares its surfaces in different views and proposed the arbitrary-view gait recognition by using partial similarity measure. Each gait was estimated through an energy cost function based on level set; shape deformation was achieved by using Laplacian deformation energy function. Wang et al. [16] designed a new gait feature representation,three-tuple gait silhouettes (TTGS), to improve the robustness of gait recognition. In [17], human gait recognition was presented based on HMM and dual discriminative observations for dynamics analysis. This method deployed both dynamics-based and model-based gait features to improve the discriminative capacity. In order to deal with the large perspective distortion in human gait recognition, Abdulsattar et al. [18] proposed an identification technique by reconstructing 3D models of walking persons. This technique treated generic Fourier descriptors (GFD) as its gait features and could deal truncated gait cycles with various lengths. Zhang et al. [20] proposed a model-based gait recognition method, which employed a five-link bipedal walking model. This method extracted the gait features from image sequences using the Metropolis-Hasting method and trained hidden Markov models based on the frequencies of trajectories. Bae and Tomizuka [21] applied the HMM models to analyze the gait phases for a patient's motions. In this method, the posterior probabilities from the HMM are used to infer the gait phases, the abnormal transition between gait phases are checked by using the transition matrix. Waugh et al. [22] proposed an individualized gait analysis approach, which was based on an adaptive periodic model of any gait signal. The proposed method learned a model of the gait cycle during online measurement, using a continuous representation that can adapt to inter and intra-personal variability by creating an individualized model. Papageorgiou et al. [23] presented a non-invasive framework for analyzing human walking patterns, which utilized a laser range finder sensor to collect the data. Besides, this solution used a synthesized HMM model for state estimation and gait recognition. Zhao et al. [24] presented an adaptive gait detection approach, which used a HMM to model human gait and employed a neural network to deal with the raw measurements. Muramatsu et al. [19] presented an arbitrary VTM model to solve the cross-view matching problem. They constructed 3D gait sequences from moving objects and generated 2D gait silhouette sequences of the training samples by projecting the 3D sequences onto the target view, then trained VTM models using the obtained 2D sequences. As a summary, these gait recognition methods mainly utilize the structural and geometric features of human walking, especially the detailed information from 3D gait models. This type of methods usually require gait video acquired simultaneously from multiple viewing angles to reconstruct the 3D gait model.

Additionally, as feature extraction is a critical preprocessing step in gait recognition [25], we will review those typical methods for feature extraction. Liu et al. [26] focused on feature selection and representation for supervised classifications.

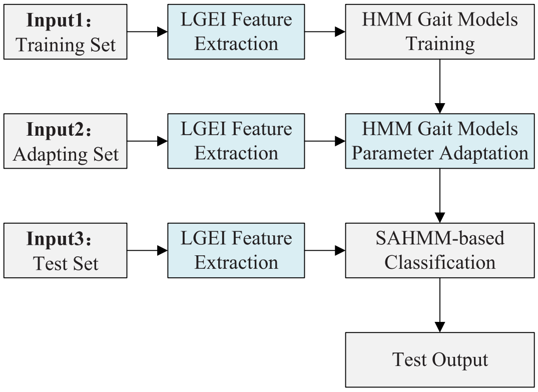


Fig. 1. The diagram of human gait recognition by using the proposed method.

A set of selection algorithms are proposed that both the selected and remaining features are relevant with the labeling. In [17], an efficient method was introduced for automatically labelling components of human body in gait silhouettes to extract body-related features. Three types of features were extracted, namely, the component area, its center and orientation. In this way, it is expected that differences in human walking can be reflected in one of the three features at least. Assuming each 3D object shares the common view surface, Tang et al. [8] presented a method for gait feature extraction by detecting such surfaces from multiple views. In the method, the gait pose was estimated by using a cost function related to level set of energy from silhouettes; body shape deformation was achieved via the function of Laplacian deformation energy associated with inpainting gait silhouettes. In [27], an automatic activity recognition method was presented, which utilized HMM to overcome the big data problem. This method incorporated personal experience into the HMM model as the priori information. He et al. [28] proposed a multichannel gait template, called period energy image (PEI), which is encoded as a view-specific feature in a latent space and then transformed from one view to another. In summary, most existing methods for extracting gait features focus on ancillary features. This process ignores the temporal association between gaits and the relationship between adjacent silhouette images of a gait. Therefore, how to mine the distinctive features of human gaits is still an unsolved problem.

Different from the existing work, in this paper, a well-designed gait representation, called LGEI, and the related feature extraction algorithm are proposed. Compared with GEI [12] and PEI [28], LGEI pays much attention to the relationship between adjacent frames and local details. Furthermore, through HMM, we present a new method for gait recognition based on SAHMM model and an incremental process to refine the parameters of each SAHMM model.

3 Methodology

Feature extraction and classifier design are the key components [29], [30] in moving object detection and recognition. In this paper, we first construct a SAHMM-based model for each person in a given gait dataset, and train them by using a new gait representation, namely LGEI. Particularly, in the training stage, we propose an incremental and adaptive algorithm to optimize the parameters of each SAHMM-based gait model. In addition, gait recognition is achieved by calculating the average probability of all models with respect to an unknown gait. The proposed method mainly includes four steps. (1) LGEI feature extraction. Gait features are extracted from the training dataset, adaptation dataset, and test dataset. (2) SAHMM training. In this step, the model is trained by using the specific training dataset. (3) Parameter adaptation. The SAHMM model is optimized by using a small amount of data from the adaptation datasets. (4) Gait classification. The SAHMM model with its optimized parameters is employed to recognize human gait from the test dataset. The diagram of gait recognition using the proposed method in this paper is shown in Fig. 1.

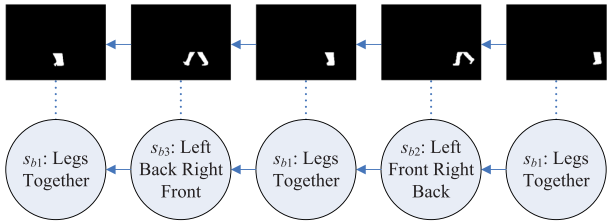


Fig. 2. A complete human gait cycle.

3.1 LGEI Feature Extraction

Feature extraction is a crucial step in human gait recognition. Usually, before gait recognition, it is necessary to detect a walking person from a given gait video. However, since the algorithms of visual object detection and tracking have been well developed [31], [32], this paper will focus on feature extraction and human gait recognition.

In accordance with the intrinsic characteristics of human walking, a gait cycle is regarded to have three basic states as shown in Fig. 2. In the basic state of Legs Together (s_{b1}), the two legs are in the same plane with the human body, which consists of three sub-states, i.e., lifting up the left foot to the beside of right leg, lifting up the right foot to the side of left leg, and standing still; the basic state of Left Front Right Back (s_{b2}) means that the left foot is on front of the right one; the basic state s_{b3} (Left Back Right Front) indicates that the right foot is on front of the left foot. Thus, a complete gait cycle is defined as $s_{b1} \rightarrow s_{b2} \rightarrow s_{b1} \rightarrow s_{b3} \rightarrow s_{b1}$; alternatively, $s_{b1} \rightarrow s_{b3} \rightarrow s_{b1} \rightarrow s_{b2} \rightarrow s_{b1}$.

After the segmentation, we extract LGEI gait features. Given the silhouette image $B_t(x,y)$ at time t in a gait cycle [12], the GEI is defined as $G(x,y) = \frac{1}{N}\sum_{t=1}^{N} B_t(x,y)$, where N is the total number of frames in a gait cycle, t is the frame indexing number, and (x,y) is the location of an image pixel. Fig. 3 shows four GEI samples, where Figs. 3a and 3b are from the same person with different angles of view; Figs. 3c and 3d are from another person with distinct view angles.

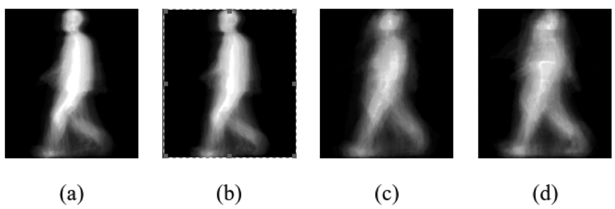


Fig. 3. GEI examples.

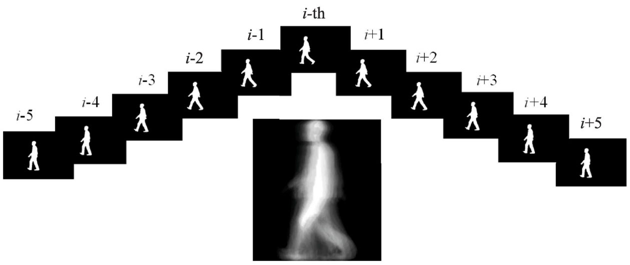


Fig. 4. Diagram of the LGEIs.

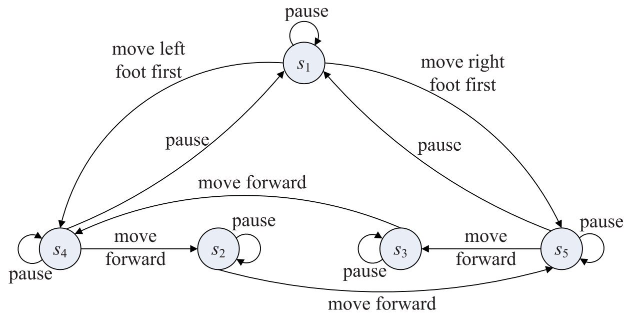


Fig. 5. State diagram of the proposed SAHMM, s_1 (Two Feet Together), s_2 (Right Through the Left Side), s_3 (Left Through the Right Side), s_4 (Left Front Right Back), s_5 (Right Front Left Back).

Generally, GEI reflects the major shape of gait silhouettes over a full cycle. GEI has three basic characteristics [33]: (1) all the silhouette images are dependent on space-normalized energy of the target individual; (2) the GEI is thought as time-normalized accumulative energy image of the target in a complete gait cycle; (3) where the pixels have bigger intensities in GEI means that the actions occur more frequently at these locations. In this paper, we refine the GEI by splitting each gait cycle into several slices and calculating LGEIs on each slice as shown in Fig. 4. Because LGEI can better preserve the details in successive frames and extract local features of human gait during view angle changing, compared with traditional GEI representation based on a complete gait cycle, LGEI can utilize the local details of human walking in each cycle to reduce the impact of viewing angle changes. The proposed feature extraction algorithm is presented in Algorithm 1.

Algorithm 1. LGEI-based Gait Feature Extraction

Input: Periodic gait dataset of a walking person.
Output: Dimension-reduced feature vector set of the walking person.
Step 1. *Find key frames.* Divide each gait cycle in the periodic gait dataset $\{c_i, i = 1, 2, \ldots, N_1\}$ and get a sequence of key frames $\{u_j, j = 1, 2, \ldots, N_2\}$, where N_1 is the number of gait cycles in the given gait dataset, N_2 is the number of key frames extracted from each gait cycle, generally, $N_2 \geq 5$. Based on the obtained experience, $N_2 = 5$ has been adopted in our experiments.
Step 2. *Calculate LGEIs for each key frame*. The calculation procedure of LGEIs within a small neighborhood of each key frame is shown in Fig. 4. It is easy to see that the use of LGEI within the neighborhood to replace a single key frame can reduce the detailed information loss.
Step 3. *Construct a feature vector set for each LGEI*. If W and H denote the width and height of each LGEI respectively, then the corresponding feature vector set $\mathcal{F} = \{f_1, f_2, \ldots, f_W\}^\tau$, where $f_k(k = 1, 2, \ldots, W)$ is an H-dimensional feature vector.
Step 4. *Reduce the dimension of feature vector sets*. In this step, we use principal component analysis (PCA) method [15] to deal with each original feature vector and obtain a new low-dimension feature set. The new feature sets will be partitioned into a training set and a test set, which will be applied to parameter estimation and performance evaluation of SAHMMs separately.

3.2 SAHMM Training

Suppose the state set of gait observation of a given person is $\{s_k, k = 1, 2, \ldots, N_3\}$, where N_3 is the number of implicit states, we construct a SAHMM model for human gait recognition $\lambda(\mathbf{A}, \mathbf{B}, \pi)$, $\mathbf{A}$ is the state transition matrix which is defined as $\mathbf{A} = (a_{ij})_{5\times 5}, a_{ij} = P(s_j, s_i)$, where a_{ij} is the probability that the implicit state is s_i at time t and the implicit state is s_j at time $t+1$. $\mathbf{B}$ is the confusion matrix defined as $\mathbf{B} = (b_{ij})_{5\times N_4}, b_{ij} = P(o_j, s_i)$, where b_{ij} is the probability at time t, the implicit state is s_i and the observed state is o_j; N_4 stands for the number of observable states. Considering that a person's walk normally is periodic and the gait cycles are relatively stable, we extract five key gaits as the implicit states of SAHMM, i.e., s_1 (Two Feet Together), s_2 (Right Foot Through the Left Side), s_3 (Left Foot Through the Right Side), s_4 (Left Front Right Back), s_5 (Right Front Left Back) as shown in Fig. 5.

By default, the gait cycle in our experiments starts from the state of standing still; in most cases, the gait begins from the state s_1 and seldom from s_4 and s_5. Thus, the state probability π is initialized, typically $\pi_1 = 0.4$, $\pi_2 = \pi_3 = 0.3$, $\pi_4 = \pi_5 = 0$.

Combined with real situation of normal walking, we set the initial values of state transition probability $\mathbf{A}$ and matrix $\mathbf{B}$ as

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0 & 1.0 \\ 0 & 0 & 0 & 1.0 & 0 \\ 0 & 1.0 & 0 & 0 & 0 \\ 0 & 0 & 1.0 & 0 & 0 \end{bmatrix} \tag{1}$$

$$\mathbf{B} = \begin{bmatrix} 1.0 & 0 & 0 & 0 & 0 \\ 0 & 1.0 & 0 & 0 & 0 \\ 0 & 0 & 1.0 & 0 & 0 \\ 0 & 0 & 0 & 1.0 & 0 \\ 0 & 0 & 0 & 0 & 1.0 \end{bmatrix}. \tag{2}$$

Given the gait feature vector set of the person m as training data, the purpose of parameter estimation for SAHMM model λ_m is to optimize the three parameters $\mathbf{A}$, $\mathbf{B}$, and $\boldsymbol{\pi}$. In human gait recognition based on image sequences, the state transition matrix $\mathbf{A}$ and the confusion matrix $\mathbf{B}$ cannot be obtained directly; hence, we use the forward-backward algorithm [34] to obtain the local optimal solution. In the algorithm, we first initialize the parameters of SAHMM, make use of the given training data to evaluate the reliability of the initialized parameters, and finally adjust the initialization parameters by minimizing the errors. Specifically, given a gait observation set $\mathcal{V}_m$, the SAHMM model λ_m and its initialization parameters, we get $\gamma_t(i) = P(q_t = s_i|\mathcal{V}_m, \lambda_m)$ and $\xi_t(i,j) = P(q_t = s_i, q_{t+1} = s_j|\mathcal{V}_m, \lambda_m)$, where $\gamma_t(i)$ is the local probability of the implicit state s_i at time t and $\xi_t(i,j)$ is the local probability of the implicit state from s_i at time t to s_j at time $t+1$. Finally, we use Eqs. (3), (4), and (5) to iteratively refine the initialization parameters and obtain a set of locally optimized parameters for $(\mathbf{A}, \mathbf{B}, \pi)$.

$$\hat{\pi}_i = \gamma_t(i) \tag{3}$$

$$\hat{a}_{ij} = \frac{\sum_{t=1}^{M} \xi_t(i,j)}{\sum_{t=1}^{T} \gamma_t(i)} \tag{4}$$

$$\hat{b}_{ij} = \frac{\sum_{t=1}^{M} \gamma_t(i)}{\sum_{t=1}^{T} \gamma_t(i)}. \tag{5}$$

In order to overcome the disadvantages of parameter adaption using maximum a posteriori(MAP) [35], we propose an incremental and adaptive method based on Cox regression analysis. By analyzing the relationship between HMM parameters, the proposed method employs a small amount of adaption data to adjust the gait model parameters effectively. Suppose the output of SAHMM-based method is mixed with Gaussian distribution, the output distributions before and after the adaption process are $\lambda = (\mu_{ij}, \sum_{ij})$ and $\widetilde{\lambda} = (\widetilde{\mu}_{ij}, \widetilde{\sum_{ij}})$. μ_{ij} and $\widetilde{\mu}_{ij}$ are the jth means of Gaussian mixed distribution of the ith state before and after the adaption process, respectively. $\sum$ and $\widetilde{\sum}$ are the covariance matrices before and after the adaption process, respectively.

3.3 Parameter Adaptation for the SAHMM

In practical applications of gait recognition, due to external reasons such as view angle of a camera, human clothing and weight bearing, there often is a mismatch between the training data and the real test data, which will lead to gait recognition declined. By using parameter adaptation which utilizes the data from a real environment to adjust the parameters of the HMM model, it is therefore possible to solve this problem. HMM parameter adaption methods usually are getting well with the maximum a posteriori estimation [35], which combine together prior knowledge and the knowledge from adaption data, then perform linear interpolation between initial parameters and adaption data to obtain final mean vector. The advantage of this method is that it makes use of prior knowledge of the model; thus, it has consistent and incremental features, especially when the adaptive data is large enough. However, the MAP-based methods are sensitive to the amount of adaptation data [36]. When the data is small, the mean after adaption will depend on the initial one.

Given an adaption dataset, the corresponding Cox regression model is defined as $\widetilde{\mu} = \mathbf{K}\mu_{ij} + \mathbf{k}_0 + \delta_{ij}$, where δ_{ij} is the residual error, $\mathbf{K}$ and $\mathbf{k}_0$ are the regression matrix and translation vector, respectively. In order to obtain $\mathbf{K}$ and $\mathbf{k}_0$, we define the objective function as $f(\mathbf{K}, \mathbf{k}_0) = \sum_{p \in \mathcal{N}} w_p \cdot \delta_p^{\tau} \cdot \delta_p$. Hence, $f(\mathbf{K}, \mathbf{k}_0) = \sum_{p \in \mathcal{N}} w_p \cdot (\widetilde{\mu} - \mathbf{K}\mu_{ij} - \mathbf{k}_0)^{\tau} \cdot (\widetilde{\mu} - \mathbf{K}\mu_{ij} - \mathbf{k}_0)$, where $\mathcal{N}$ is the nearest neighbor set of λ_m before the adaption process; w_p is the weight factor that is defined as $w_p = \exp(-d_p^2)$ and $d_p^2 = \sum_{k=1}^{K} \frac{(\mu_p^k - \mu_{ij}^k)^2}{\beta_{ij}^k}$, where β_{ij}^k is the kth diagonal element of covariance matrix $\sum_{ij}$, namely, $\sum_{ij} = diag(\beta_{ij}^1, \ldots, \beta_{ij}^K)$.

From the definition, when the distance is smaller, the corresponding weight factor w_p will be larger. Thus, it has a great impact on the regression model. Meanwhile, if there is no corresponding adaption data for a mean vector, we only need to find the regression class and utilize the corresponding matrix to conduct the adaption transformation. For a small capacity of gait recognition based on SAHMM, one leaf node in its regression tree represents a single component and each node on the high layer stands for a set of components with similar distance; meanwhile, the root node contains all the mixture components. When a SAHMM has multiple mixed components, the leaf nodes come along with the basic classes based on the initial clustering, each of the basic classes comprehends to a set of components with similar distance. The experimental results show that the proposed method is effective for gait recognition with only a small amount of adaption data.

3.4 SAHMM-Based Human Gait Recognition

Given a gait observation set $\{v_k | k = 1, 2, \ldots, N_4\}$, the observation set $\mathbf{O}_k = \{o_{k_1}, o_{k_2}, \ldots, o_{k_T}\}$ corresponds to v_k, and the set of all samples in gait dataset, gait recognition is converted to the problem of SAHMM evaluation. Thus, we only need to calculate the average probability $\bar{P}_m = \frac{1}{n}\sum_{k=1}^{N} P(\mathbf{O}_k|\lambda_m)$, where each SAHMM model outputs the results $\{\bar{P}_m\}$ given the observation set. We make use of the forward algorithm to calculate the probability that each SAHMM model generates the observation set $\mathbf{O} = \{\mathbf{O}_k | k = 1, 2, \ldots, N_4\}$. The human gait recognition based on SAHMM is described as Algorithm 2.

Algorithm 2. Human Gait Recognition using SAHMM with Parameter Adaptation

Input: The feature vectors from the training set and test set.
Output: The model number n.
Step 1. *Construct SAHMM models*. We construct a SAHMM-based gait model $\lambda_m(\mathbf{A},\mathbf{B},\pi)$ for each person in current gait dataset and initialize parameters for all models as Eqs. (1) and (2).
Step 2. *Estimate parameters of each gait models*. Based on the feature vectors extracted from the training set, we employ the forward-backward algorithm [34] to estimate the local optimized parameters of all gait models.
Step 3. *Refine the results of parameter estimation*. In this step, we refine the parameters of each gait model by using the proposed incremental and adaptive method. Besides, since this step encompasses parameter optimization for the SAHMM model, which is relatively complex, the related operations are carried out offline.
Step 4. *Calculate the average probability of all models*. The local probability $\gamma_t(j)$ of each hidden state is computed by using recursions as shown in Eq. (6). For each observation set $\{v_k | k = 1, 2, \ldots, N_4\}$ of the test set, the corresponding probability of each SAHMM model equals to the sum of all the local probabilities at time T as shown in Eqs. (6) and (7).

$$\gamma_t(j) = \begin{cases} \pi(j) b_{jk_t} & t = 1 \\ \sum_{i=1}^{5} \gamma_{t-1}(j) a_{ij} b_{jk_{t-1}} & t = 2, 3, \ldots, T \end{cases} \tag{6}$$

$$\bar{P}_m = \frac{1}{n}\sum_{k=1}^{N} P(\mathbf{O}_k|\lambda_m) = \frac{1}{n}\sum_{k=1}^{N}\sum_{j=1}^{5} \gamma_T(j) \tag{7}$$

Step 5. *Implement gait recognition*. In this step, we sort the average probabilities $\{\bar{P}_m\}$ related to each gait model and get the model number n with the maximum average probability.

Fig. 6. The 11 examples with various view angles in CASIA Dataset B.

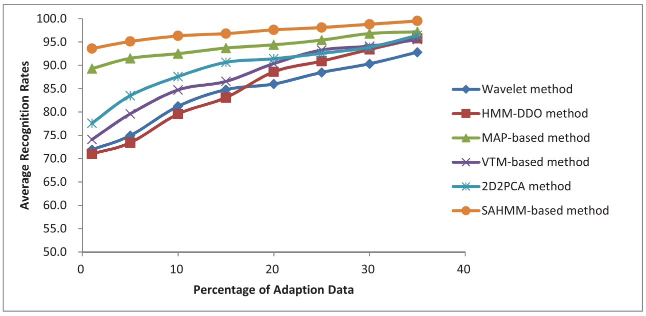


Fig. 7. The average recognition rates of different methods in Experiment I.

4 Experimental Results

Based on the CASIA Dataset B [37], OU-ISIR Large Population Dataset [38], and OU-ISIR Multi-View Large Population Dataset(OU-MVLP) [39], we designed and implemented five experiments for evaluating the proposed gait recognition methods. Experiment I and Experiment II are to evaluate gait recognition under the conditions of cross view and dressing or bag carrying. The CASIA Dataset B contains the videos of 124 working persons including 11 views as shown in Fig. 6, distributed from 0 to 180° under three kinds of conditions, i.e., normal walking, wearing coats and carrying bags. Experiment III is conducted to verify the generalization capability of the proposed methods by using the OU-ISIR Large Population Dataset which contains gait images of 4,016 walking persons, including 12 image sequences, 4 sequences for each of the three directions, i.e., 0° (parallel), 45°, and 90°(perpendicular) to the image plane. Experiment IV is conducted on OU-MVLP dataset to further evaluate the proposed method with multi-view large-scale gait dataset. The OU-MVLP dataset consists of 10,307 subjects (5,114 males and 5,193 females with various ages, ranging from 2 to 87 years) from 14 view angles, ranging from $0° to$-90°, from $180° to$-270°. Gait images with $1{,}280 \times 980$ pixels at 25 FPS are captured by seven network cameras (Cam1-7) placed at intervals of 15-deg azimuth angles along a quarter of a circle whose center coincides with the center of the walking course. The OU-MVLP dataset was split into two disjoint subsets, i.e., the training set (taged as OU-MVLP-A) and the test set (taged as OU-MVLP-B). The final experiment is conducted for comparing two gait feature representations: the proposed LGEI and traditional GEI.

Average recognition rate and standard deviation are taken into consideration in the course of evaluations, which reflect the stability of relevant gait recognition methods. These experiments utilize subsets of the CASIA dataset to compare the proposed method with other three methods. The experimental datasets were obtained by using random sampling.

4.1 Experiment I. Comparisons of Experimental Results by using Partial Data in the CASIA Dataset B

This experiment makes use of the gait data with normal walking condition from CASIA Dataset B. We totally construct 8 training sets, which consist of all the data with a view angle of 90° plus randomly selected data from other views with proportions 1, 5, 10, 15, 20, 25, 30, and 35 percent. For adaptive training-based methods, we utilize the data with a view angle of 90° for preliminary training, and then use the selected data from other view angles for adaptive training. For others, we directly apply the constructed sets for training.

Figs. 7 and 8 show the comparisons by using the wavelet-based method [14], the HMM-DDO method [17], the MAP-based adaptive method [35], the VTM-based method [19], the $(2D)^2$PCA method [15], and the SAHMM-based method, where the horizontal axis is percentage of adaption data, vertical axis in Fig. 7 indicates average recognition rates, and the vertical axis in Fig. 8 presents the standard deviations of recognition rates.

From Figs. 7 and 8, we see that the proposed method in this paper has higher average recognition rates and less standard deviations. This owes to the use of adaptive parameter optimization, which easily leads to the mismatch between the training data and the test data. Furthermore, Figs. 7 and 8 show that with the increase of the proportions of adaption data from the test data, the average recognition rates of the four methods have all been improved. However, the methods have a wide spectrum of sensitivity to the adaption data; the standard deviations of recognition rates of the methods are different. The wavelet-based and HMM-DDO methods have not adaptive process when the size of adaption data is small, the training data with a view angle of 90° plays a decisive role in the process of parameter estimation and the recognition rates are very low; but when the adaption data is increased to more than 20 percent, which is very close to the normal amount of training data, the recognition rates will be increased sharply. Pertaining to the methods after an adaptive parameter optimization process, the recognition rates are already very high in the case of adaption data with the increase of adopted training data, the curves are rather smooth.

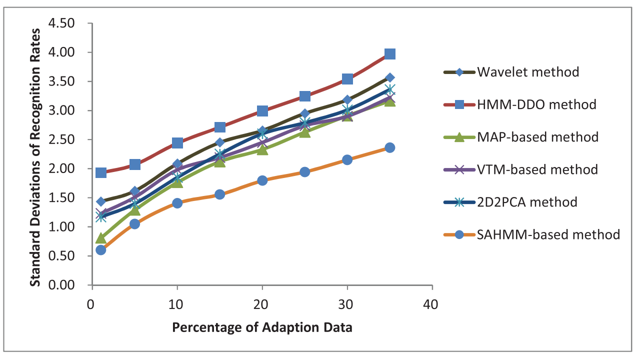


Fig. 8. The standard deviations of different methods in Experiment I.

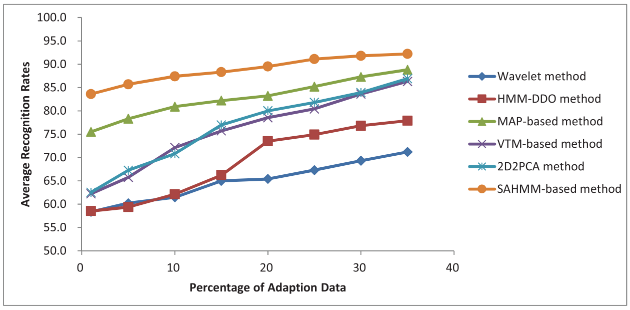


Fig. 9. The average recognition rates of different methods in Experiment II.

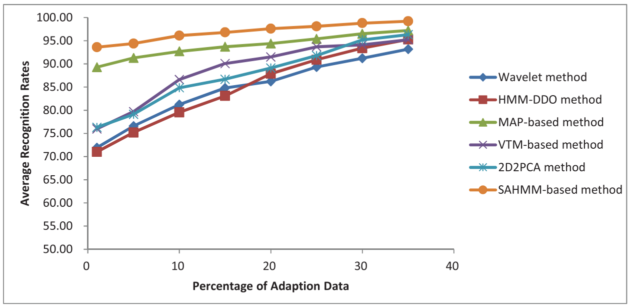


Fig. 11. The average recognition rates of different methods in Experiment III.

4.2 Experiment II. Comparisons of Experimental Results by using All Data of the CASIA Dataset B

This experiment adopted all the gait samples in CASIA Dataset B including the gait data under walking conditions of wearing coats and carrying bags. Similar to Experiment I, we constructed 8 types of training sets from all gait data with a view angle of 90° and randomly selected gait data from the rest with proportions of 1, 5, 10, 15, 20, 25, 30, and 35 percent. Similar to Experiment 1, for the methods with an adaptive training process, the samples with a view angle of 90° are used for preliminary training, and then the selected samples from other views are employed for adaptive training; while for other methods, we directly use the constructed training sets for training. Figs. 9 and 10 show the experimental results with the horizontal axis for representing the percentages of total adaption data and vertical axis in Fig. 9 for displaying the average recognition rates, the vertical axis in Fig. 10 stands for the standard deviations of the obtained recognition rates.

In Figs. 9 and 10, compared to Experiment I, the average recognition rates of the four methods all have a significant decline due to the increase of training data which has very complicated reasons such as wearing coats and carrying bags. But our algorithm has higher average recognition precision, especially when the adaption data is less. From Fig. 10, we see that with the increase of adaption data, the standard deviations of gait recognition of the four methods have been significantly increased; but the standard deviations after an adaptive parameter optimization are relatively low. The reason is that the increase of abnormal gait data, such as wearing a coat and carrying a bag, results in the decrease of recognition rates. On the other hand, the methods with a process of adaptive parameter optimization reduce the imbalance between the training data and the test data, the corresponding standard deviations of recognition rates are rather small.

4.3 Experiment III. Comparison of Experimental Results by using OU-ISIR Large Population Dataset

In this experiment, the effectiveness of our proposed method is assessed against the OU-ISIR Large Population Dataset, namely OULP-C1V2. This dataset comprises of two subsets: A and B. A is a set of two sequences per subject and B is a set consisting of one sequence per subject. In addition, each of the main subsets is further split into 5 subsets based on the observation angles of 55°, 65°, 75°, and 85° including all four angles. The dataset consists of over 4,000 individuals walking on the ground surrounded by two cameras at 30 FPS with the resolution 640×480.

During the training phase, we utilize eight different training datasets including all data in subset A plus randomly selected data from the subset B with proportions of 1, 5, 10, 15, 20, 25, 30, and 35 percent. Then, the constructed training datasets are directly used to train all the comparative methods, except for the two methods with an adaptive training process, i.e., the MAP-based adaptive method [35] and the SAHMM-based method, which are trained twice using all the data from the subset A and the selected data from the subset B, respectively. The relevant results are shown in Figs. 11 and 12, where the horizontal axis represents the percentage of the total adaption data, the vertical axis in Fig. 11 stands for the average recognition rates, the vertical axis in Fig. 12 shows the standard deviations of these recognition rates as well.

Fig. 11 shows the MAP-based adaptive method, the proposed method in this paper has obvious advantages with

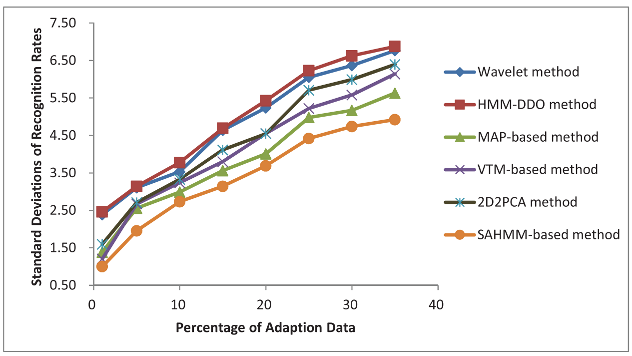


Fig. 10. The standard deviations of different methods in Experiment II.

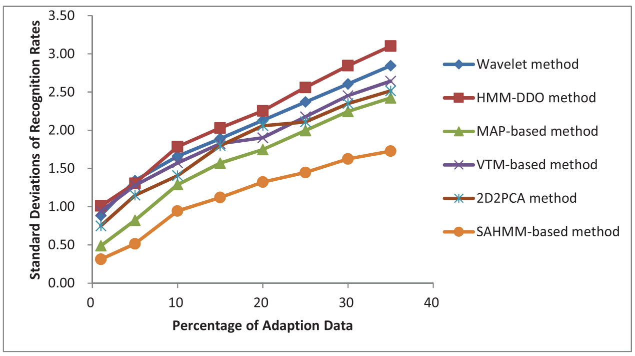


Fig. 12. The standard deviations of different methods in Experiment III.

TABLE 1
Comparison With State-of-the-Art Gait Recognition Methods on MVLP Dataset in Terms of Correct Recognition Rates (CRR) and Standard Deviations (SD)

METHOD	CRR1(%)	CRR2(%)	CRR3(%)	CRR4(%)	CRR5(%)	CRR6(%)	CRR7(%)	CRR8(%)	AVERAGE CRR(%)	SD
$(2D)^2$PCA [15]	49.25	50.73	52.76	54.87	57.61	60.49	62.91	65.74	68.63	4.20
RTN [41]	61.03	62.86	65.38	67.99	71.39	74.96	77.96	81.47	85.05	2.52
HBPL [40]	56.46	58.15	60.48	62.90	66.04	69.35	72.12	75.37	78.68	3.86
CVGR-EL [42]	58.55	60.31	62.72	65.23	68.49	71.91	74.79	78.16	81.59	3.18
MAP [35]	69.06	73.20	76.86	79.94	82.34	83.16	83.58	83.99	84.41	1.65
Our Method	74.18	78.63	82.56	85.86	88.44	89.33	89.77	90.22	90.67	1.47

CCR1,CCR2,. . ., CCR8 are the CCRs corresponding to the training sets that include selected data from OU-MVLP-B with proportions of 1, 5 percent,. . ., 35 percent, respectively.

regard to the average recognition rates, compared with the wavelet-based and HMM-DDO methods. The reason is that the process of adaptive parameter optimization partially solves the mismatch problem between the training data and the real test data. As shown in Fig. 12, multiple adaptive methods have different sensitivities to the adaption data from the test dataset; hence, the stability of these recognition rates is obvious in the repeated experiments. On the whole, the recognition rates of these methods with an adaptive parameter optimization process are much stable; thus, the corresponding standard deviations are rather small.

The experimental results demonstrate the effectiveness of this proposed algorithm. This owns to the well-designed algorithm for gait feature extraction and partially benefits from the parameter adaptation optimization in the proposed SAHMM-based method for human gait recognition.

4.4 Experiment IV. Comparison of Experimental Results by Using OU-ISIR Multi-View Large Population Dataset

To further evaluate the proposed method with multi-view large-scale gait dataset, we conduct Experiment IV on OU-ISIR Multi-view Large Population Dataset. Different from other experiments, in addition to the MAP-based adaptive method [35], the $(2D)^2$PCA method [15], we added three latest methods:the HBPL-based method [40], the RTN-based method [41], the CVGR-EL method [42] in this experiment. During the training phase, we construct eight training datasets by combining the OU-MVLP-A and randomly selected data from the OU-MVLP-B with proportions of 1, 5, 10, 15, 20, 25, 30, and 35 percent, respectively. Then, we train the MAP-based adaptive method [35] and our method twice by using the OU-MVLP-A and randomly selected data from the OU-MVLP-B. Because other methods have not an adaptive training process, we directly train them by using the eight training datasets we have collected.

The correct recognition rates and standard deviation of the proposed method are compared with those of existing state-of-the-art methods in Table 1. It is obvious that as the proportion of data from OU-MVLP-B increases, i.e., from CCR1 to CCR8, all involved methods perform better. The reason is that the number of training data is gradually increased from CCR1 to CCR8, the corresponding model training is much adequate.

In addition, from CCR1 to CCR8, the CCR increase of the two methods with an adaptive training process is more obvious because the adaptive training process can further optimize the model parameters with a small number of gait samples from the real environment. Furthermore, the proposed method has better recognition rates than the MAP-based method. This is reflected in a significant improvement in the CCR values and a decrease in the corresponding SD values.

Finally, compared with Experiment I to Experiment III, we can find that the overall recognition rate of this experiment is low, the main reason is that the dataset used in this experiment has two characteristics: cross-view and large capacity.

4.5 Experiment V. Comparisons of Two Gait Representations: GEI and LGEI

In this section, we compare the performance of two gait features, i.e., GEI and LGEI, based on CASIA Dataset B and OU-ISIR LP Dataset through the two sub-experiments. GEI and LGEI are used as the inputs of the proposed gait recognition method so as to compare the average correct recognition rates.

In the first sub-experiment, like Experiment II, we used all the gait samples from CASIA dataset B to construct 8 test cases. For each test, we constructed the training sets utilizing all gait samples with 90° view angle and randomly selected samples from the rest with proportions of 1, 5, 10, 15, 20, 25, 30, and 35 percent. During the training process, we use the data with 90° view angle for primary training and use other data for further adaptive training. The average recognition rates are shown in Table 2. We see from Table 2, by using LGEI representation, we have a better recognition result than that by using GEI representation, particularly when the percentage of adaptation data is smaller. The main reason is that LGEI can correctly reflect most of the local features of human gaits, thus reduce the effect from view angle changing.

In the second sub-experiment, similar to Experiment III, we designed eight test cases using the OU-ISIR Large Population Dataset. The training set in each case is collected from all data in subset A plus randomly selected data from subset B with proportions of 1, 5, 10, 15, 20, 25, 30, and 35 percent. The related experimental results are shown in Table 3.

TABLE 2
Comparisons of Average Correct Recognition Rates (%) of Two Gait Representations Based on CASIA Dataset B

Percentage of Adaption Data	1%	5%	10%	15%	20%	25%	30%	35%
GEI	69.5	71.2	74.0	76.2	79.1	80.8	83.7	84.2
LGEI	83.6	85.7	87.4	88.3	89.5	91.1	91.8	92.2

TABLE 3
Comparisons of Average Correct Recognition Rates (%) of Two Gait Representations on OU-ISIR LP Dataset

Percentage of Adaption Data	1%	5%	10%	15%	20%	25%	30%	35%
GEI	85.7	86.2	88.5	89.1	91.6	92.6	93.2	95.7
LGEI	93.6	94.4	96.1	96.8	97.6	98.1	98.8	99.2

Table 3 shows that even with more large-scale dataset, the LGEI gait presentation can benefit the classification in terms of correct recognition rates. The results in Table 3 indicate that the correction recognition rates achieved by using LGEI and GEI are better than that in the first sub-experiment. The main reason is that the gait samples in CASIA Dataset B are captured under more complex walking conditions than those in OU-ISIR LP Dataset.

5 Conclusion

In this paper, a well-designed feature representation and a feature extraction algorithm are presented. Then, a human gait recognition method based on SAHMM is proposed. We construct the observed state set from each gait cycle and train the SAHMM model. In addition, in order to improve our gait recognition, we design and implement an adaptive algorithm to automatically adjust the parameters of SAHMM model.

To the best of our knowledge, this is the first time we work for gait recognition using the SAHMM-based algorithm; our contributions are that we, (1) present the feature extraction algorithm based on LGEI which utilizes important details of gait features in each gait cycle; (2) propose the human gait recognition method based on SAHMM model which makes use of a parameter adaptation process to optimize the parameters of each gait model; (3) promote the gait recognition based on the CASIA Dataset B and OU-ISIR Large Population Dataset, demonstrate that our method can work well by using the gait datasets. The limitation of the proposed algorithm is that, in each gait image sequence, the changes of walking direction should not be too extensive. In future, we will work on how to reduce or weaken the restricted conditions and how to improve the algorithms for gait recognition so as to match real needs.

Acknowledgments

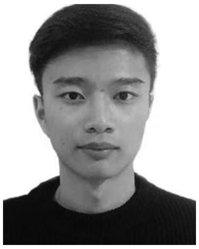
We thank reviewers for their constructive suggestions and valuable comments on our work. This work was supported by the NSFC of China under Grant No.61303146 and a scholarship of the China Scholarship Council (CSC).

References

[1] P. Connor and A. Ross, "Biometric recognition by gait: A survey of modalities and features," *Comput. Vis. Image Understanding*, vol. 167, pp. 1–27, 2018.
[2] C. Wan, L. Wang, and V. V. Phoha, "A survey on gait recognition," *ACM Comput. Surveys*, vol. 51, no. 5, pp. 1–35, 2018.
[3] Z. Wu, Y. Huang, L. Wang, X. Wang, and T. Tan, "A comprehensive study on cross-view gait based human identification with deep CNNs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 02, pp. 209–226, Feb. 2017.
[4] N. Takemura, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi, "On input/output architectures for convolutional neural network-based cross-view gait recognition," *IEEE Trans. Circuits Systems for Video Technol.*, vol. 29, no. 9, pp. 2708–2719, Sep. 2019.
[5] T. Connie, M. Goh, and A. Teoh, "A Grassmannian approach to address view change problem in gait recognition," *IEEE Trans. Cybern.*, vol. 47, no. 06, pp. 1395–1408, Jun. 2017.
[6] J. Luo, J. Tang, and T. Tjahjadi, "Robust arbitrary view gait recognition based on parametric 3D human body reconstruction and virtual posture synthesis," *Pattern Recognit.*, vol. 60, pp. 361–377, 2016.
[7] X. Wang and W. Q. Yan, "Human gait recognition based on frame-by-frame gait energy images and convolutional long short term memory," *Int. J. Neural Syst.*, 2019, doi: 10.1142/S0129065719500278.
[8] J. Tang, J. Luo, and T. Tjahjadi, "Robust arbitrary-view gait recognition based on 3D partial similarity matching," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 7–23, Jan. 2017.
[9] P. Chung and C. Liu, "A daily behavior enabled hidden Markov model for human behavior understanding," *Pattern Recognit.*, vol. 41, no. 5, pp. 1572–1580, 2008.
[10] D. Feng, C. Bao, W. Kleijn, and W. Bastiaan, "Sparse hidden Markov models for speech enhancement in non-stationary noise environments," *IEEE-ACM Trans. Audio, Speech, Language Process.*, vol. 23, no. 11, pp. 1973–1987, Nov. 2015.
[11] X. Liu and T. Chen, "Video-based face recognition using adaptive hidden Markov models," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 207–212.
[12] J. Han and B. Bhanu, "Individual recognition using gait energy image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 02, pp. 316–323, Feb. 2006.
[13] Y. Guan, C. Li, and F. Roli, "On reducing the effect of covariate factors in gait recognition: A classifier ensemble method," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 07, pp. 1521–1529, Jul. 2015.
[14] M. Islam, M. Islam, M. Hossain, A. Ferworn, and M. Molla, "Subband entropy-based features for clothing invariant human gait recognition," *Adv. Robot.*, vol. 31, no. 08, pp. 519–530, 2017.
[15] X. Wang, J. Wang, and K. Yan, "Gait recognition based on Gabor wavelets and $(2D)^2$PCA," *Multimedia Tools Appl.*, vol. 77, no. 10, pp. 12 545–12 561, 2018.
[16] X. Wang, J. Zhang, and W. Yan, "Gait recognition using multichannel convolution neural networks," *Neural Comput. Appl.*, 2019, doi: 10.1007/s00521-019-04524-y.
[17] N. Boulgouris and X. Huang, "Gait recognition using HMMs and dual discriminative observations for sub-dynamics analysis," *IEEE Trans. Image Process.*, vol. 22, no. 09, pp. 3636–3647, Sep. 2013.
[18] F. Abdulsattar and J. Carter, "Performance analysis of gait recognition with large perspective distortion," in *Proc. IEEE Int. Conf. Identity Security Behav. Anal.*, 2016, pp. 1–6.
[19] D. Muramatsu, A. Shiraishi, Y. Makihara, M. Z. Uddin, and Y. Yagi, "Gait-based person recognition using arbitrary view transformation model," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 140–154, Jan. 2015.
[20] R. Zhang, C. Vogler, and D. Metaxas, "Human gait recognition at sagittal plane," *Image Vision Comput.*, vol. 25, no. 3, pp. 321–330, 2007.
[21] J. Bae and M. Tomizuka, "Gait phase analysis based on a hidden Markov model," *Mechatronics*, vol. 21, no. 6, pp. 961–970, 2011.
[22] J. L. S. Waugh *et al.*, "Online learning of gait models from older adult data," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 4, pp. 733–742, Apr. 2019.
[23] X. S. Papageorgiou, G. Chalvatzaki, C. S. Tzafestas, and P. Maragos, "Hidden Markov modeling of human normal gait using laser range finder for a mobility assistance robot," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2014, pp. 482–487.
[24] H. Zhao *et al.*, "Adaptive gait detection based on foot-mounted inertial sensors and multi-sensor fusion," *Inf. Fusion*, vol. 52, pp. 157–166, 2019.
[25] K. Yan, L. Ma, Y. Dai, W. Shen, Z. Ji, and D. Xie, "Cost-sensitive and sequential feature selection for chiller fault detection and diagnosis," *Int. J. Refrigeration-Revue Internationale Du Froid*, vol. 86, pp. 401–409, 2018.
[26] Y. Liu, F. Tang, and Z. Zeng, "Feature selection based on dependency margin," *IEEE Trans. Cybern.*, vol. 45, no. 6, pp. 1209–1221, Jun. 2015.
[27] Z. Li *et al.*, "An adaptive hidden Markov model for activity recognition based on a wearable multi-sensor device," *J. Med. Syst.*, vol. 39, no. 5, Mar. 2015, Art. no. 57.
[28] Y. He, J. Zhang, H. Shan, and L. Wang, "Multi-task GANs for view-specific feature learning in gait recognition," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 1, pp. 102–113, Jan. 2019.
[29] K. Yan, Z. Ji, H. Lu, J. Huang, W. Shen, and Y. Xue, "Fast and accurate classification of time series data using extended ELM: Application in fault diagnosis of air handling units," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 49, no. 7, pp. 1349–1356, Jul. 2019.

[30] Y. Liu, R. Xiong, and Y. Li, "Robust and accurate multiple-camera pose estimation toward robotic applications," *Int. J. Adv. Robotic Syst.*, vol. 11, pp. 1–8, 2014.
[31] J. Kwon and K. Lee, "Adaptive visual tracking with minimum uncertainty gap estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 01, pp. 18–31, Jan. 2017.
[32] K. Yan, C. Zhong, Z. Ji, and J. Huang, "Semi-supervised learning for early detection and diagnosis of various air handling unit faults," *Energy Buildings*, vol. 181, pp. 75–83, 2018.
[33] P. Theekhanont, W. Kurutach, and S. Miguet, "Gait recognition using GEI and pattern trace transform," in *Proc. Int. Symp. Inf. Technol. Medicine Educ.*, 2012, pp. 936–940.
[34] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statistical Soc.*, vol. 39, no. 01, pp. 1–38, 1977.
[35] S. Yin and W. Liang, "Experimental study of discriminative adaptive training and MLLR for automatic pronunciation evaluation," *Tsinghua Sci. Technol.*, vol. 16, no. 2, pp. 189–193, 2011.
[36] K. Yan, Z. Ji, and W. Shen, "Online fault detection methods for chillers combining extended kalman filter and recursive one-class SVM," *Neurocomputing*, vol. 228, no. SI, pp. 205–212, 2017.
[37] S. Yu, D. Tan, and T. Tan, "A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition," in *Proc. Int. Conf. Pattern Recognit.*, 2006, pp. 441–444.
[38] H. Iwama, M. Okumura, Y. Makihara, and Y. Yagi, "The OU-ISIR gait database comprising the large population dataset and performance evaluation of gait recognition," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 5, pp. 1511–1521, Oct. 2012.
[39] N. Takemura, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi, "Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition," *IPSJ Trans. Comput. Vision Appl.*, vol. 10, no. 4, pp. 1–14, 2018.
[40] I. Rida, X. Jiang, and G. L. Marcialis, "Human body part selection by group lasso of motion for model-free gait recognition," *IEEE Signal Process. Lett.*, vol. 23, no. 1, pp. 154–158, 2016.
[41] S. Tong, Y. Fu, and H. Ling, "Cross-view gait recognition based on a restrictive triplet network," *Pattern Recognit. Letters*, vol. 125, pp. 212–219, 2019.
[42] X. Wang and W. Q. Yan, "Cross-view gait recognition through ensemble learning," *Neural Comput. Appl.*, 2019, doi: 10.1007/s00521-019-04256-z.

Xiuhui Wang received the PhD degree in computer science and technology, State Key Laboratory of CAD&CG from Zhejiang University China, in 2007. From 2018 to 2019, he was an academic researcher with the Auckland University of Technology, Auckland, New Zealand. He is currently an associate professor with the China Jiliang University, Hangzhou, China. His research interests include computer vision, pattern recognition, and computer graphics.



Shiling Feng is currently working toward the master's degree in computer science and technology at the China Jiliang University. His research interests include computer vision and pattern recognition.

Wei Qi Yan received the PhD degree in computer engineering from the Chinese Academy of Sciences, in 2001. He is currently an associate professor with the Auckland University of Technology, New Zealand. His research interests include intelligent surveillance, computer vision, pattern recognition, and computer graphics. He is a senior member of the IEEE.