Neural Computing and Applications (2020) 32:7275–7287
https://doi.org/10.1007/s00521-019-04256-z

ORIGINAL ARTICLE

Cross-view gait recognition through ensemble learning

Xiuhui Wang[1] · Wei Qi Yan[2]

Received: 10 January 2019 / Accepted: 9 May 2019 / Published online: 16 May 2019

Abstract

Gait has been well known as an unobtrusive promising biometric to identify a person from a distance. However, the effectiveness of silhouette-based approaches in gait recognition is diluted due to variations of view angles. In this paper, we put forward a novel and effective method of gait recognition: cross-view gait recognition based on ensemble learning. The proposed method greatly enhances the effectiveness and reduces the sensitivity of gait recognition under various view angles conditions. Furthermore, in this paper we will introduce a novel algorithm based on ensemble learning for combining several gait learners together, which utilizes a well-designed gait feature based on area average distance. Through experimental evaluations on the well-known CASIA gait database and OU-ISIR gait database, our paper demonstrates the advantages of the proposed method in comparison with others. The contribution of this research work is to resolve the multiview angles problem of gait recognition through assembling several gait learners.

Keywords Gait recognition · Cross-view · Ensemble learning · Gait classification

1 Introduction

Gait is composed of complicated human actions integrated movements of body parts and joints together. As an effective biometric [1, 2], human gait has a great deal of advantages over classical biometrics, such as human face, fingerprint, iris. (1) Unobtrusiveness. Most of the existing biometrics require direct contact, specific postures or fixed emotions for pattern classifications, while gait recognition generally works in a natural way which does not need special notifications for identifying persons. (2) Gait images and videos are usually captured from a very far distance. Unlike face recognition, gait recognition utilizes vision techniques to identify a person with a very low resolution. Thus, gait recognition is a simple, attractive, and effective method. (3) Gait images and videos are easily taken using normal cameras like the one set forth in a mobile phone. Even with low-quality images, gait recognition algorithms can still work [3].

However, gait recognition remains a challenging problem in real applications, namely most of the existing gait recognition algorithms only work well under the best condition of image and video acquisition [4]. Correspondingly, the performance of gait recognition algorithms is dimmed due to varying illumination, view angles, human clothing, etc. In this paper, a novel identification approach of cross-view gait recognition based on ensemble learning (CVGR-EL) is proposed. The CVGR-EL method takes use of a novel gait representation, i.e., gait features based on area average distance (AAD), which enhances the effectiveness and reduces the sensitivity to variations of various view angles conditions. In this paper, we will develop a novel algorithm to ensemble several learners and achieve gait recognition.

The contributions of this research work are summarized as follows: (1) A well-designed gait feature representation. Based on area average distance, we propose a novel feature representation and corresponding feature extraction method that makes the best use of intrinsic characteristics of human gaits and improves gait recognition rate. (2) A novel view-

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant Nos. 61303146 and 61602431 as well as a scholarship from the China Scholarship Council (CSC).

✉ Xiuhui Wang
wangxiuhui@cjlu.edu.cn

Wei Qi Yan
wyan@aut.ac.nz

1 China Jiliang University, No. 258, Xueyuan Street, Hangzhou 310018, China

2 Auckland University of Technology, No. 2-14, Wakefiled Street, Auckland 1010, New Zealand

dependent base-learners training using the proposed gait feature representation. We draw random gait training sets from the extracted AAD-based gait features for training a set of HMM learners. (3) Gait classification based on ensemble learning. To combine multiple base learners, we present a novel voting-like method.

This paper is structured as follows. The state-of-the-art technologies of gait recognition are presented in Sect. 2. Next, the feature extraction and classification algorithms for gait recognition are described in Sect. 3. In Sect. 4, our experimental results by using CASIA gait dataset [5] and OU-ISIR gait database [6] are shown and discussed. Finally, in Sect. 5 we will conclude the contributions of this paper and envision our future work.

2 Related work

As emerging biometrics, human gait has attracted more and more attention of scientists [1, 2, 4, 5] in digital forensics and computer vision. Compared to those traditional biometric technologies, such as fingerprint and face recognition, gait recognition has to confront more challenges and is currently in its infancy. One of the challenges is how to overcome the problem of angular variety between the directions of the camera pointed in and the target person, the camera may be held with hand or fixed by a tripod in this case. Another challenge in gait recognition is the problem caused by clothes and belongings of the walking person, such as wearing a heavy coat or carrying with a very big bag. These problems affect the recognition rate of gait recognition. The focus of existing work is mainly on feature selection, feature extraction, and recognition rates. According to gait features, most of the recent gait recognition methods are roughly grouped into three categories: static-feature-based, dynamic-feature-based, and hybrid-feature-based methods.

Approaches of gait recognition based on static-gait features [7–12] directly utilize geometric information of human body from picture sequences and generally have higher computational cost. Moustakas et al. [7] proposed an efficient framework for augmenting gait recognition with soft biometric information. It was confined to simple and static gait features such as user height and stride length. Huang and Boulgouris [8] presented a gait recognition system which used the shifted energy image and the gait structural profile. Two gait features are robust to appearance variations such as clothing and carrying conditions, but are still sensitive to angle variations of view. In [9], Procrustes shape analysis was used for gait feature description and similarity measurement; a higher-order shape configuration is proposed for gait shape description. This method tackled the problem caused by walking speed change. Rida et al. [10] proposed a gait recognition method for selecting the most discriminative part of the human body based on group Lasso of motion. This method can significantly improve the recognition rate in the case of clothing variation. Theekhanont et al. [11] presented a gait recognition method by using Gait Energy Image (GEI) and pattern trace transform. This method is simple and easy to be implemented, but its recognition rate needs to be improved. Furthermore, a Gabor wavelets-based gait recognition algorithm was proposed [12], which employs two-dimensional principal component analysis ($(2D)^2$PCA) method to reduce the dimension of feature space. In a word, these static features of gait recognition are generally confronting with changing appearances of a walking person under a diversity of conditions such as walking speed, clothing and belongings of the walker, shooting angle of the camera, etc.

On the other hand, those gait recognition methods based on dynamic features have also drawn a great attention. The methods based on dynamic features [13–21] take temporal characteristics of gait into consideration, which reflects the differences of human posture, motion amplitude and walking velocity. Zhang et al. [13] proposed a method, namely super-resolution with manifold sampling and back-projection, to learn the dynamic gait features, which is much efficient with low-resolution cases. In [14], a sparse bilinear discriminant analysis based on dynamic gait features using discriminant subspace learning was proposed. Muramatsu et al. [15] proposed a gait recognition algorithm that can achieve a high recognition rate, which only focused on the case of view issue and omitted the other factors. In [16], an arbitrary view transformation model was presented, which can match a pair of gait features from an arbitrary view. However, its performance is not always so good in verification tasks without score normalization. Sruti and Tardi [17] proposed a two-phase view-invariant gait recognition method, which is robust to variation in clothing and carried condition. Nonetheless, this method requires the availability of closely matching test view with the training dataset. Hu et al. [18] presented an incremental framework based on optical flow, including dynamics learning, pattern retrieval and recognition. This method improved the usability of gait recognition in video surveillance applications, but suffered from view angle variations. Connie et al. [19] proposed a method which can generate virtual views to compensate the view difference in query and reference sets. The method offers a bridge between conventional methods and manifold-matrix-based approaches for gait recognition problems. Muramatsu et al. [20] presented a view transformation model which incorporates a score normalization framework with quality measures that encode the degree of the bias. The limitations of this method are that the recognition rate is little improved in several angle of view, an additional training dataset for parameter evaluation is required. In [21], a Siamese neural network-based gait recognition framework was developed to extract robust and discriminative gait features. Because its architecture of distance metric learning can well solve the verification problem of gait recognition, this method can achieve a better recognition rate in the case of moderate variations of view angle.

Because both static features and dynamic features of human gaits are merely a subset of the entire feature set and cannot reflect inherent characteristics of human gait globally which eventually affect the recognition rate of gait recognition partially. Therefore, to improve the final correct classification rate (CCR), hybrid-feature-based methods [22–28] have provided a promising way. Boulgouris and Huang [22] proposed a gait recognition method, which used the holistic and model-based gait features to enhance the discriminatory capacity. Islam et al. [23] presented a wavelet-based feature extraction method for gait recognition, and a template matching-based approach is employed in the classification process. Aggarwal and Vishwakarma [24] proposed a covariance conscious approach to deal with variations in clothing and carrying conditions. In this method, Zernike moment invariants were used, gait features were extracted from spatial distribution. Wang et al. [25] described a gait recognition method by combining static and dynamic biometrics. This method used both Procrustes shape information and joint-angle trajectories of lower limbs, which improved the recognition rate. Cuntoor et al. [26] presented a gait recognition method, which employed different gait features, and this method used both probabilistic and non-probabilistic techniques to match the features. Nevertheless, because this method utilized the swing of hands, it was sensitive to variations of carrying conditions. Veres et al. [27] proposed a gait recognition method based on fusion of dynamic and static features which aimed to deal with the problem of gait recognition when time-dependent covariates are involved. The method can improve the recognition rate with time-dependent covariance, but it generally involved reducing the dimensionality of datasets and obtained different sets of features by using feature refining algorithms. In [28], a gait recognition method was proposed to analyze silhouette boundary and further convert it into associated 1D signals. This special idea reduces the computational cost of the subsequent processes. In addition, with the successful applications of Deep Learning techniques in image representation and processing [29, 30], some researchers tried to solve the problem of gait recognition by using deep neural networks (DNN) [2, 31]. Although in theory, as long as the number of layers in a DNN is enough, a good set of features can always be extracted, the training costs of DNN are very high and the extracted features are not intuitive.

The existing work has shown that gait recognition is a sophisticated problem, which is dependent on feature selection, feature extraction, learner design and other factors. To the best of our knowledge, there are quite few of comprehensive studies [32–34] for gait recognition by taking ensemble learning into consideration though a lot of research work has implemented multiple learners in this field. Thus, in this paper, we will present a new solution for cross-view gait recognition based on ensemble learning (CVGR-EL). Compared to previous approaches, the new solution will employ a novel gait representation, namely AAD-based gait features. Secondly, we use a scaling-factor-based method to train a set of view-dependent HMM gait learners. Finally, we propose a novel gait recognition based on ensemble learning and design three experiments to test the advantages of our method on CASIA gait database and OU-ISIR gait database.

3 Our proposed methods

Feature selection and feature extraction are fundamental problems in gait recognition. In the CVGR-EL method, a well-designed gait feature is selected and used in our learners. The proposed method mainly includes five steps: (1) Half gait-cycle segmentation to split a gait video by analyzing differences of the entire region of a walker's human body. (2) AAD-based gait feature extraction. In this step, we extract a new gait feature based on the distance between two regions; (3) View-dependent HMM models training. A set of HMM gait learners are constructed and trained according to different views; (4) Gait classification based on ensemble learning. A set of view-dependent HMM learners are used to classify gaits in a test set and the results are combined. The flowchart of our algorithm is shown in Fig. 1, where a set of AAD-based gait features from the view k is used for training the HMM learner, the outcome of learner k is used as the input of HMM classifications.

As the preprocess of feature extraction, automatically gait-cycle segmentation is the initial step of gait recognition. Taking into account of computational complexity of the segmentation algorithm and diversity of multiple views of human gaits, we extract each half gait cycle by counting the number of pixel changes in the region of the entire body of a walker. Observational results show that in the process of human walking, the area of the walker's body alters regularly as shown in Fig. 2. When two legs of a gait separate to the maximum distance, no matter from which view, the area of gait region reaches its local maxima. We therefore define the boundary of the half gait cycle as the

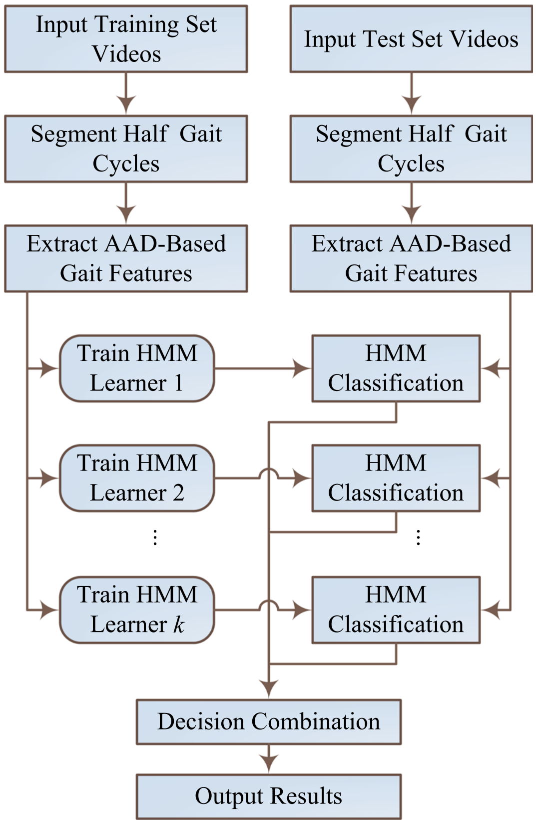


Fig. 1 Flowchart of the CVGR-EL algorithm

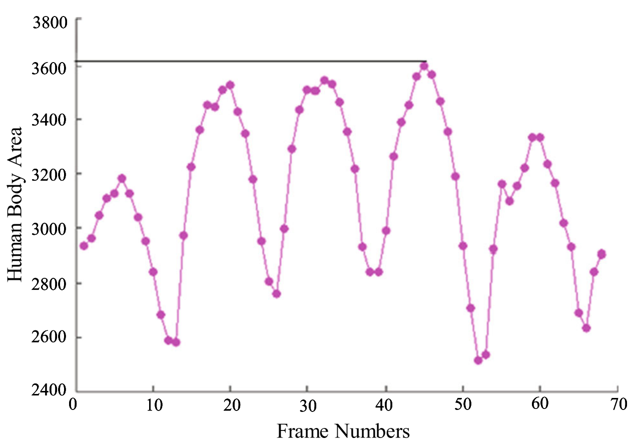


Fig. 2 The area changes of a worker's body in consecutive video frames

threshold so as to extract a series of half-cycle gait when the area reaches its maxima. However, when the angle is less than 10 degrees, especially close to degree 0, there is a possibility that the automatic segmentation fails; therefore, we need to adjust the parameters.

3.1 AAD-based gait features extraction

There are two main methods to extract static-gait features: one is based on the silhouettes of walkers' body; the other is based on the region. Both of these two methods have their advantages and disadvantages. The first one better reflects the detailed changes of human gait; however, it requires highly accurate operations to extract silhouettes; therefore, the silhouette segmentation needs more time. In this paper, a novel method of static-gait feature extraction based on the average distance is proposed, which is able to make the best use of walkers' gait and improve gait recognition significantly. After foreground detection and binarization processing [1], a pixel $P(x, y)$ of the silhouette image I is expressed as,

$$P(x,y) = \begin{cases} 1 & (x,y) \in H \\ 0 & \text{others} \end{cases}, \tag{1}$$

where H is a collection of pixels located at the body region. Based on the silhouette of the segmented region of the body, we calculate barycentric coordinates of the gait area as shown in Fig. 3. The centroid of this gait region is calculated as defined in Eqs. (2) and (3).

$$x_c = \frac{1}{n}\sum_{x,y} x \cdot P(x,y), \tag{2}$$

$$y_c = \frac{1}{n}\sum_{x,y} y \cdot P(x,y), \tag{3}$$

where n represents the number of pixels in gait area which is defined as,

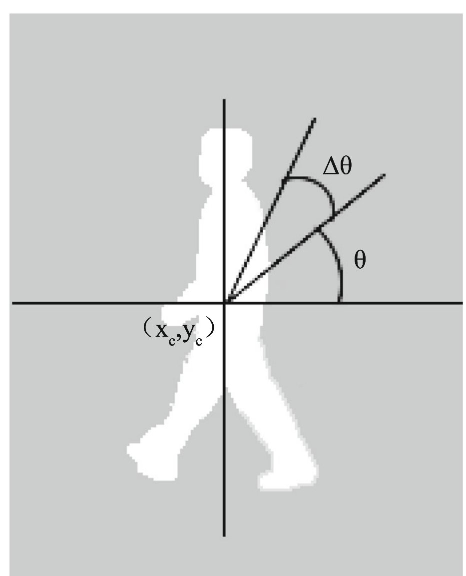


Fig. 3 The centroid of a gait area

$$n = \sum_{x,y} P(x, y). \tag{4}$$

Finally, we set up a two-dimensional Cartesian coordinate system with the centroid as the origin and define the static-feature variable of an image as,

$$\mu(\theta) = \frac{1}{n_{H_\theta}} \sum_{x,y} P(x, y)\sqrt{(x - x_c)^2 + (y - y_c)^2} \tag{5}$$

where H_θ represents the pixel set in the range $(\theta, \theta + \Delta\theta)$, n_{H_θ} is the number of pixels in H_θ, $\mu(\theta)$ is the average distance of all the pixels in H_θ to the barycentric center (x_c, y_c) of the region of human body. In this paper, we calculate the average distance between all the pixels within the angle range, the region center of the human body is shown in Fig. 3. The sum of these distances in a gait region is defined as,

$$D(\theta) = \int_{\theta}^{\theta+\Delta\theta} \int_{0}^{r_\theta} r^2 \mathrm{d}r\mathrm{d}\theta. \tag{6}$$

The area of a region of human body with an angle of θ is defined as,

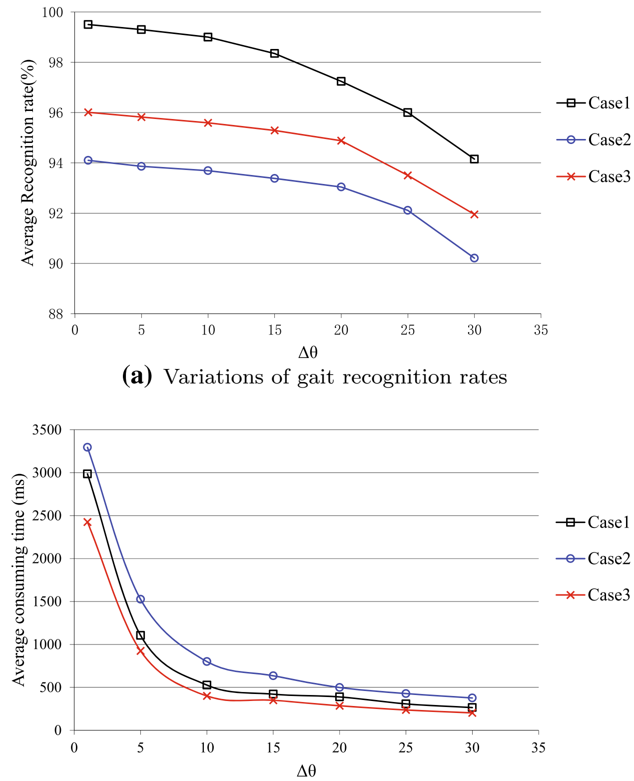
$$V(\theta) = \int_{\theta}^{\theta+\Delta\theta} \int_{0}^{r_\theta} r\mathrm{d}r\mathrm{d}\theta. \tag{7}$$

Based on Eqs. (6) and (7), we derive the average distance from all pixels in the given region of the centroid of the body $d(\theta)$ as shown in Eq. (8) and construct the static feature vectors that will be used as the input of each HMM gait model.

$$d(\theta) = \frac{D(\theta)}{V(\theta)}. \tag{8}$$

By utilizing the average distance of gait area as a feature representation, we propose a novel method of static-feature extraction for gait recognition. The scale of the final gait vector is decided on the parameter $\Delta\theta$ as shown in Eqs. (6)–(8). From Fig. 3, we see that the value range of $\Delta\theta$ is [1, 360]. Even when we set $\Delta\theta$ to 1°, the final gait vector only has 360 elements, which is far less than that in traditional GEI-based methods. Furthermore, based on the experimental result, as shown in Fig. 4, we usually set $\Delta\theta$ to 10°, which will reduce computational complexity of the proposed method greatly.

Meanwhile, based on experimental experience, we select the angular parameter $\Delta\theta$ as 10° and get a 36-dimension feature vector. When $\Delta\theta$ adopts a larger value, the recognition rate will be significantly reduced; on the other hand, when $\Delta\theta$ is smaller than 10°, the recognition rate will be lifted very small, but the corresponding computational time will be increased as shown in Fig. 4. With regard to Case 1 of our experiment, we only use the normal data in CASIA gait dataset B; for Case 2, we adopt all the data in CASIA dataset B; for Case 3, we utilize all the data in CASIA dataset A. Hereinafter, Case 1, Case 2 and Case 3 are a set of experiments we will detail them later in Sect. 4.



(a) Variations of gait recognition rates

(b) Variations with the computational time

Fig. 4 Variations of recognition rates and computational complexity with the changes of $\Delta\theta$

3.2 View-dependent training of HMM learners

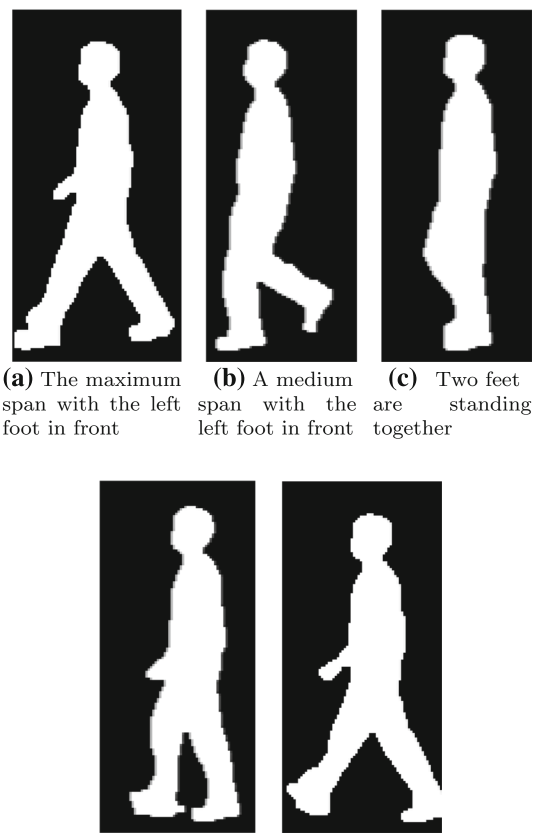
In this section, we will discuss the training process of base learners. We ensemble several view-dependent learners which are trained by using gait data from different views of the same individual separately. In the proposed cross-view gait recognition, we use HMM as base learners which yield a ranking together with a score function. HMM is one of the time series methods which have been extensively used in speech, pedestrian and human behavior recognition. However, when HMM is employed to a complex problem, like gait recognition, it is still a promising solution in parameter optimization and model training.

To simulate the natural way of knowledge acquisition, HMM models are utilized to classify human gaits based on static features. The forward–backward algorithm is a special case of Expectation Maximization (EM) [35], which is based on dynamic programming to improve model parameters. In the iterations, the forward variable α and the backward variable β are used in the forward–backward algorithm, α rapidly approaches to 0 as time goes by; meanwhile, β quickly tends to 0. In the calculation of α, β will have very small intermediate values, which will result in the problems of numerical underflow. Thus, we apply scaling factors to the traditional forward–backward algorithm of HMM, which makes the training process of HMM more robust. In addition, we propose a discriminant distance-based method to initialize matrix B of the HMM model.

In a half gait cycle, there are two feet moving simultaneously in the alternative states: lift right leg, lift left foot, so on and so forth. We extract key frames from each half gait cycle automatically, which will be used to define the HMM model of each walker as shown in Fig. 5. A gait model based on HMM is denoted as,

$$\lambda = (N, M, \pi, \mathbf{A}, \mathbf{B}), \tag{9}$$

where N is the number of hidden states, M is the length of observation sequence, π is the initial state probability matrix, A is the state transition probability matrix, and B is the transition matrix between visible and latent states.



(a) The maximum span with the left foot in front **(b)** A medium span with the left foot in front **(c)** Two feet are standing together

(d) A medium span with the right foot in front **(e)** The maximum span with the right foot in front

Fig. 5 Five key states in a half gait cycle

To select the optimal number of hidden states, many existing methods use the trial and error strategy, which involves computing a score. In our work, a k-means-based algorithm was used to decide the optimal number of hidden states [36]. In the HMM model, the initial value of π and transition matrix A have little influence on the sequential model as mentioned in the literature [37]. Based on our experience, in human normal walking procedure, the speed is roughly constant; the five key states, as shown in Fig. 5, have similarly approximate probabilities, so we simply set the initial value of π as Eq. (10); besides, considering the continuity of the five key states, we set the initial value of A as Eq. (11), instead of a simple diagonal matrix.

$$\pi = [0.2, 0.2, 0.2, 0.2, 0.2] \tag{10}$$

$$\mathbf{A} = \begin{bmatrix} 0.5 & 0.3 & 0.07 & 0.03 & 0.1 \\ 0.1 & 0.5 & 0.3 & 0.07 & 0.03 \\ 0.03 & 0.1 & 0.5 & 0.3 & 0.07 \\ 0.07 & 0.03 & 0.1 & 0.5 & 0.3 \\ 0.3 & 0.07 & 0.03 & 0.1 & 0.5 \end{bmatrix} \tag{11}$$

Initialization of the matrix $\mathbf{B}$ is a bit tough, which will greatly influence the training result of gait model. We present a discriminant distance-based method to create initial matrix $\mathbf{B}$ as,

$$\mathbf{B} = \left[\alpha e^{\beta D(x(t), e_n)}\right], \tag{12}$$

where α and β are constants, $D(x(t), e_n)$ is the distance between the observation vector at time t and the nth latent vector e_n. In order to find out matrix B, we need to obtain the state sequence $e_1, e_2, \ldots, e_n$ so that we use the k-means algorithm to cluster the given gait sequences and set each cluster center as the value of ε. Given the number of samples in clustering ϕ_j as K_j, the corresponding average value $\overline{m}_j$ is calculated as,

$$\overline{m}_j = \frac{1}{K_j} \sum_{i=1}^{K_j} l_i, \tag{13}$$

For each cluster, we obtain the sum of Euclidean distances between the corresponding center and all the samples in the class, i.e., errant sum of the squares. Then, the discriminant distance to initialize matrix B is calculated by adding up all the errant sums of each cluster together. When Eq. (14) reaches the minima or the clustering centers do not change in the iterative process, we get the optimized clustering results and therefore construct the target state sequence $e_1, e_2, \ldots, e_n$ by using,

$$S_e = \sum_{n}^{t=1} \sum_{l \in \phi_j} \left\| l - m_j \right\|. \tag{14}$$

After having established HMM gait models for all gaits and initialized them, we present an improved forward–backward algorithm based on scaling factors to train the HMM model. The new forward variables are defined as,

$$\alpha_1(i) = \pi(i) b_{ik_1}, \tag{15}$$

$$\alpha_1^*(i) = \frac{\alpha_1(i)}{\sum_{i=1}^{N} \alpha_1(i)} = \frac{\alpha_1(i)}{\Phi_1}, \tag{16}$$

$$\widetilde{\alpha}_{t+1}(j) = b_{jk_{t+1}} \sum_{i=1}^{N} \alpha_t^*(i) a_{ij}, \tag{17}$$

$$\alpha_{t+1}^*(j) = \frac{\widetilde{\alpha}_{t+1}(j)}{\sum_{j=1}^{N} \widetilde{\alpha}_{t+1}(j)} = \frac{\widetilde{\alpha}_{t+1}(j)}{\Phi_{t+1}}, \tag{18}$$

where $1 \le i \le N$, $1 \le j \le N$ and $t = 1, 2, \ldots, T-1$.

Accordingly, the new backward variables are,

$$\beta_T(i) = 1, 1 \le i \le N, \tag{19}$$

$$\beta_T^*(i) = \frac{1}{N}, 1 \le i \le N, \tag{20}$$

$$\widetilde{\beta}_t(i) = \sum_{j=1}^{N} a_{ij} b_{jk_{t+1}} \beta_{t+1}^*(j), 1 \le t \le T-1, \tag{21}$$

$$\widetilde{\beta}_t(i) = \widetilde{\beta}_t(i) / \Phi_t. \tag{22}$$

In addition, when calculating the probability $P(O|\lambda)$ generated by a sequence of observations in a HMM model, the method to eliminate the scaling factor is also improved. In the optimization of each forward variable, scaling factors are used to optimize the process, we deduce,

$$\alpha_t^*(i) = \alpha_t(i) / \Phi_1 \Phi_2 \ldots \Phi_t, \tag{23}$$

$$\Phi_t = \sum_{j=1}^{N} \widetilde{\alpha}_t(j) = \sum_{j=1}^{N} \alpha_t(j) / \Phi_1 \Phi_2 \ldots \Phi_{t-1}. \tag{24}$$

Then, we get,

$$\sum_{j=1}^{N} \alpha_t(j) = \Phi_1 \Phi_2 \ldots \Phi_t. \tag{25}$$

Finally, $P(O|\lambda)$ is calculated by,

$$P(O|\lambda) = \sum_{j=1}^{N} \alpha_T(j) = \Phi_1 \Phi_2 \ldots \Phi_T, \tag{26}$$

In the improved algorithm, $P(O|\lambda)$ needs to be calculated according to Eq. (26). Because both the numerator and denominator contain the common factor to be canceled, the equation can be directly simplified. Meanwhile, when more than one observation sequence of the training HMM models is adapted, the evaluations of HMM are redefined as,

$$\overline{\pi}_i = \frac{\sum_{l=1}^{L} \gamma_1^{(l)}(i)}{L}, \tag{27}$$

$$\overline{\alpha}_{ij} = \frac{\sum_{l=1}^{L} \sum_{t=1}^{T_l - 1} \xi_t^{(l)}(i,j)}{\sum_{l=1}^{L} \sum_{t=1}^{T_l - 1} \gamma_t^{(l)}(i)}, \tag{28}$$

$$\overline{b}_{jk} = \frac{\sum_{l=1}^{L} \sum_{t=1, o_t = k}^{T_l} \gamma_t^{(l)}(j)}{\sum_{l=1}^{L} \sum_{t=1}^{T_l} \gamma_t^{(l)}(j)}, \tag{29}$$

where $1 \le i \le N$, $1 \le j \le N$, and $1 \le k \le M$.

According to the steps of iterations, we gradually refine the parameters of HMM gait models until the parameters of the three matrices in the model are basically unchanged or the value of $P(O|\lambda)$ is maximized, we get the optimized parameters. When the sum of squares of differences of all the matrix parameters between two adjacent iterations is lower than a given threshold, we think these matrices to be substantially unchanged. In our experiments, this threshold was set to 0.001. In a nutshell, by introducing scaling factors into the forward–backward algorithm, we thus avoid the problems of numerical underflow and effectively estimate the parameters of each gait model.

3.3 Gait recognition based on ensemble learning

Ensemble learning is to enhance recognition rates through combining a set of view-dependent HMM learners together. Compared with existing ensemble learning methods, the outstanding feature of our method is the exploration of intrinsic characteristics of human gaits through a well-designed gait feature representation and the decision combination process by using Bayesian inference. Our aim is to ensemble n base learners $L_i (i = 1, 2, \ldots, n)$ with an additional combination learner C. The base learners L_i are working with the original AAD-based gait features, while the combination learner C is based on the output of the base learners L_i. There are three main problems to be solved in gait recognition based on ensemble learning: (1) extract well-designed gait features for training process; (2) select valid base learners and train them; (3) choose a combination strategy to ensemble these base learners into a strong learner. Till now, we have solved the first two problems by extracting AAD-based gait features and used them to train HMM learners. This section will focus on how to ensemble these base learners into a combination one. Formally, we set the ranking returned by the ith base learner as:

$$R_{H_i} = ((c_1, v_1), (c_2, v_2), \ldots, (c_M, v_M)), \tag{30}$$

where M is the number of the ranks from base learner L_i, $c_j (j = 1, 2, \ldots, M)$ is the code related to the input gait, and

v_M is the value of corresponding score function on rank j. Following the framework proposed in [38], we construct the score function based on a probability distribution that is estimated by transforming each score to the empirical distribution. The rank is listed in ascending order according to the scores. Finally, the combination learner is fed with the output of the base learners and yields a refined ranking.

The proposed gait recognition algorithm based on ensemble learning is presented as Algorithm 1.

Algorithm 1 Gait recognition based on ensemble learning

Input: Training video and test video
Output: Results of gait recognition
Step 1. Segment input training and test videos into half gait cycles, group them into several categories according to their view angles. The number of groups is determined by the capacity of gait database, type of gait data, and quantity of each gait data.
Step 2. Extract the AAD-based gait features as described in Sect. 3.1
Step 3. Train view-dependent HMM gait learners. Each HMM learner is trained using a different feature set extracted from special view angles, and the detailed training process is described in Sect. 3.2.
Step 4. Gait classification using the trained HMM models. The forward algorithm is employed to classify the feature vectors using the well-trained HMM model, the corresponding class confidence is calculated.
Step 5. Decision combination. In this step, we use Bayesian inference to derive the joint probability of each HMM learner. Assuming that the event, which the gait of the jth individual has been imported to the base learner L_i, is $E_j(j = 1, 2, \ldots, N)$, N is the total number of individuals in our gait database, O_k denotes the fact that the leaner L_i classifies the input gait to the kth individual, then we can obtain the probability of gait recognition by using Bayesian theorem,

$$P(E_j|O_k) = \frac{P(E_j)P(O_k|E_j))}{\sum_{l=1}^{N} P(E_l)P(O_k|E_l))}, \tag{31}$$

Given a gait E_j, the classification is corrected with probability p. Then, we obtain

$$P(O_k|E_j) = \begin{cases} p & k = j \\ \dfrac{1-p}{N-1} & k \neq j \end{cases} \tag{32}$$

Finally, after estimating the matching threshold [39], for making an ensemble decision, we compute the joint probability,

$$P\left(E_j|\bigcap_{l=1}^{N} O_k^l\right) = \frac{P(E_j)\left(\prod_{l-1}^{N} P(O_k^l|E_j)\right)}{\left(\prod_{l=1}^{N} P(O_k^l|E_j)\right)\sum_{l=1}^{N} P(E_l)}. \tag{33}$$

In summary, in order to improve the precision of cross-view gait recognition, we propose a novel algorithm of gait recognition based on ensemble learning, which combines a set of view-dependent HMM learners into a strong learner and conducts gait recognition.

The algorithm proposed in this paper is a sequential structure, in which the step with the highest time complexity is the third step, i.e., the EM-based HMM training. The EM is an algorithm with low computational complexity and low storage overhead, which can be calculated with very small computational resources [40]. As linear algebra is used in the EM algorithm, its complexity is $O(m \cdot n^3)$, where m is the number of iterations and n is the number of parameters. Furthermore, as shown in Eqs. (15–22), the process of adding scaling factors does not change the time complexity of the original algorithm.

4 Experimental results

In this section, three experiments have been carried out based on the well-known CASIA gait database [28, 41] and the OU-ISIR gait database [6]. Because these gait databases are publicly available with limited access restrictions and widely used in a great deal of the state-of-the-art work, we used them to design comparative experiments so as to evaluate the identification performances of the proposed algorithms. The first experiment was conducted on the CASIA gait dataset A [28] (former NLPR Gait Dataset). The purpose of this experiment is to test the performance of the proposed method and compare the recognition rates of the proposed method with the state-of-the-art gait recognition techniques. The second experiment was performed by using the CASIA gait dataset B [41] that is composed of gait videos from a variety of view angles and complicated interference conditions, such as wearing different fashion of clothes, carrying a bag, etc. The purpose of the second experiment was to analyze the impact of interference conditions against the recognition rate of the proposed method. The final experiment was conducted using one of the datasets in OU-ISIR gait database, which is the largest gait dataset at present. The purpose of the third experiment was to generalize the proposed methods. Besides, the learners in all the experiments are working under a closed set for identification. We trained them with the data from all individuals; in the test phase, each gait is classified as one of the known individuals in the corresponding gait database.

Throughout the experiments, the performance of each related method was measured by using the Correct Classification Rates (CCR),

$$CCR = \frac{TP + TN}{TP + FP + TN + FN}, \quad (34)$$

where TP, TN, FP, and FN are the number of true positive, true negative, false positive, and false negative. To visually compare the recognition performances of different methods, cumulative match characteristic (CMC) is employed as another evaluation criterion, which is regarded as a thorough comparison because it is based on the variation of ranks. When calculated the CCR, the rank-1 recognition rate is used; a correct recognition occurs only when a sample in the test set is the best matching one from the training set.

The performance of this proposed CVGR-EL method is compared to the existing methods [11, 19, 20, 23–25]. The method in [11] is based on static gait features and the methods [19, 20] mainly used dynamic gait features; meanwhile, the methods in [23–25] utilized both types of gait features. Besides, the work [19, 20] was specially designed to handle cross-view gait recognition, while the others [11, 23–25] are general gait recognition methods. When we implemented and ran the experiments for these methods, the default parameters reported in the original papers were deployed.

4.1 Experiments based on CASIA dataset A

In this experiment, the effectiveness of this proposed method is assessed against the CASIA gait dataset A, which consists of 20 subjects and each subject has 12 image sequences, 4 sequences for each of the three view angles, i.e., parallel (marked as 0-ISs), 45° (marked as 45-ISs) and 90° (marked as 90-ISs) to the image plane. Three base HMM learners are constructed and trained using training sets from three different views separately. We design three test cases (TCs) to study the influence of view angles on CCR as shown in Table 1. Because more gait features could be easily extracted from the 90-ISs, both Test Case 1 and Test Case 2 are in use of more samples from the 90-ISs to form the training set. As the complementary, the final test takes use of samples which are eventually selected from all types of image sequences. In all of the test cases, the split samples were randomly repeated 1000 times, and the mean and standard deviation (SD) are presented in Table 2.

The CCRs for each test set are compared with the existing state-of-the-art methods in Table 2. It is obvious that the proposed CVGR-EL method achieves perfect CCR with Test Case 1 and Test Case 2. This is attributed to the static-gait features which make best use of gait intrinsic characteristics. More importantly, the proposed method performs extremely well in the presence of Test Case 3, where it achieves 95.5% recognition rate. This is 7.7% better than AESI+ZNK [24] which achieved a recognition rate of 88.1%. Figure 6 shows CMC curves for the average recognition rate of the three test cases. Among all the seven curves, the proposed method CVGR-EL shows a better recognition rate. Thus, the proposed method performs well in comparison with those existing methods.

4.2 Experiments on CASIA dataset B

The CASIA dataset B provides clear separation of view angles and other external factors, such as clothing, carrying, which enables our evaluations due to other effects, such as view and variations on gait recognition. The CASIA gait dataset B consists of 124 subjects captured from 11 different angles. Three variations, namely view angle, clothing, and carrying condition changes, are considered. The view angles, ranging from 0 degree to 180 degrees, are separated by an interval of 18°. Corresponding to each view angle, there are ten videos for each subject with six videos having the subjects: walking under normal condition (marked as *N*-videos), two videos with the subjects: wearing coats (marked as C-videos), and two videos with the subjects: carrying bags (marked as B-videos).

In this experiment, we divided the gait data with 11 different angels into six sets, i.e., (0°,180°), (18°, 172°), (36°, 144°), (54°, 126°), (72°, 108°), and the final dataset (90°). Data from each set are used to train one base HMM learner respectively; further to study the variations of CCRs with a range of training data, we construct our own training sets and test sets to form four test cases (TCs) as shown in Table 3. In Case 1, we only used *N* videos and mainly evaluate the influence of variations in view angles. Both Case 2 and Case 3 do not take variations into account in view angles and focus on other external conditions, e.g., with or without wearing coats, whether or not carrying bags, etc. The final case takes both the variations of view

Table 1 The composition of the training and test sets in experiment A

	Training set	Test set
TC1	All the 90-ISs and 50% of the 0-ISs	The remaining image sequences
TC2	All the 90-ISs and 50% of the 45-ISs	The remaining image sequences
TC3	50% of all types of images sequences	The remaining image sequences

Table 2 Comparison of CCRs with the state-of-the-art gait recognition methods on casia dataset A

Method	TC1 (%)	SD1	TC2 (%)	SD2	TC3 (%)	SD3	Average (%)	Total SD
GEI+PTT [11]	81.5	0.5	82.5	0.43	82.4	0.5	82.13	0.55
GVC [19]	83.2	0.47	84.1	0.46	83.9	0.36	83.73	0.47
VTM [20]	85.2	0.35	86	0.4	85.5	0.33	85.57	0.4
wEnI [23]	86	0.15	86.3	0.11	85.9	0.2	86.07	0.21
AESI+ZNK [24]	87.3	0.43	88.2	0.42	88.1	0.46	87.87	0.49
PSA [25]	90.4	0.87	92.2	0.79	91.7	0.83	91.42	0.93
CVGR-EL	95.5	0.35	96.3	0.33	95.8	0.32	95.87	0.4

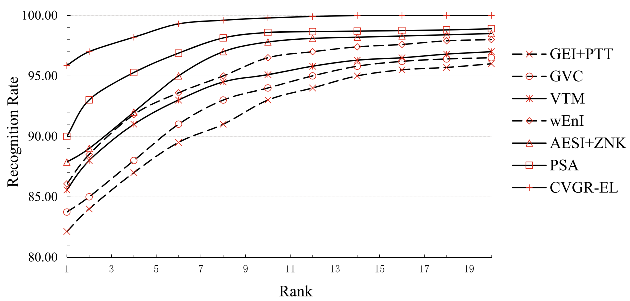


Fig. 6 Comparisons using CMC curves based on the CASIA Dataset A

angles and the other external conditions into account. Similar to Experiment A, the split samples were randomly repeated 1000 times in the test, the mean and standard deviation are presented in Table 4.

The CCRs by using seven different methods are shown in Table 4 which shows that the proposed method performs superior in the presence of Test Case 1. The recognition rate achieved by the proposed algorithm is 96.1%, which is 5.8% better than the existing state-of-the-art methods; AESI+ZNK [24] achieved 90.3% recognition rate. The average recognition rate reached 92.7%, which is 8% better than the existing state-of-the-art, wEnI [23]. Figure 7 shows CMC curves for the average recognition rate of the four test cases. From Fig. 7, it is apparent that the recognition rate achieved by the proposed method is the highest one. This is understandable because the proposed method ensembles two base learners and conduct gait recognition.

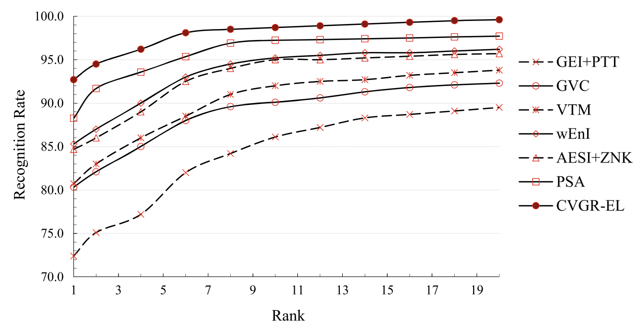


Fig. 7 Comparison using CMC curves based on the CASIA dataset B

Table 3 The composition of the training and test sets in experiment B

	Training set	Test set
TC1	50% of *N* videos with 90 degrees	The remaining *N* videos
TC2	50% of *N* videos with 90 degrees	The remaining videos with 90 degrees
TC3	50% of all videos with 90 degrees	The remaining videos with 90 degrees
TC4	50% of all videos with 90 degrees	The remaining videos

Table 4 Comparison of CCRs with the state-of-the-art gait recognition methods on CASIA dataset B

Method	TC1 (%)	SD1	TC2 (%)	SD2	TC3 (%)	SD3	TC4 (%)	SD4	Average (%)	Total SD
GEI+PTT [11]	83.1	9.8	72.5	9.85	74.8	9.9	59.2	9.51	72.4	9.91
GVC [19]	85.7	4.38	79.0	4.43	81.5	4.36	75.1	3.91	80.3	4.45
VTM [20]	88.9	5.59	77.4	5.59	79.9	5.54	76.7	5.07	80.7	5.62
wEnI [23]	86.7	1.64	84.6	1.70	86.7	1.63	83.2	1.02	85.3	1.71
AESI+ZNK [24]	90.3	4.15	83.1	4.19	85.0	4.13	80.3	4.01	84.7	4.22
PSA [25]	93.2	4.02	88.5	4.00	89.6	3.99	83.4	3.48	88.7	4.05
CVGR-EL	96.1	2.60	91.3	2.74	93.6	2.65	89.8	2.64	92.7	2.75

Table 5 The composition of the training and test sets in experiment C

	Training set	Test set
TC1	50% of the samples in Seq00	The remaining image sequences in Seq00
TC2	50% of the samples in Seq01	The remaining image sequences in Seq01
TC3	50% of the samples in both Seq00 and Seq01	The remaining image sequences in Seq00 and Seq01

Table 6 Comparison of CCRs with the state-of-the-art gait recognition methods on OU-ISIR large population dataset

Method	TC1 (%)	SD1	TC2 (%)	SD2	TC3 (%)	SD3	Average (%)	Total SD
GEI+PTT [11]	75.6	2.79	76.1	2.83	80.8	2.86	77.5	2.9
GVC [19]	77.1	2.66	79.5	2.72	82.7	2.71	79.8	2.8
VTM [20]	79.9	2.11	77.3	2.11	81.6	2.16	79.6	2.2
wEnI [23]	83.2	2.18	82.6	2.21	86.9	2.23	83.6	2.3
AESI+ZNK [24]	83.1	2.36	85.9	2.3	87.9	2.38	85.6	2.4
PSA [25]	90.6	1.49	91	1.47	93.5	1.51	91.7	1.6
CVGR-EL	95.7	1.3	97.2	1.29	98.5	1.33	97.1	1.4

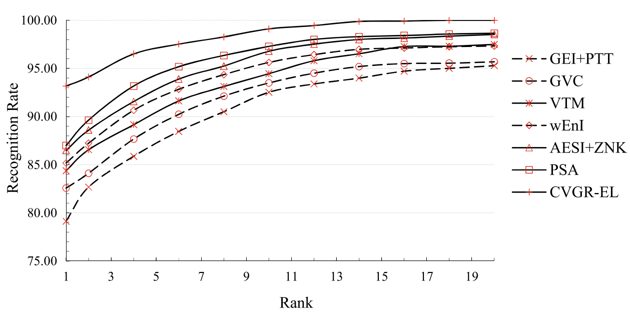


Fig. 8 Comparisons using CMC curves based on the OU-ISIR large population dataset

4.3 Experiments on OU-ISIR large population dataset

In this experiment, the effectiveness of the proposed method is assessed against the OU-ISIR Large Population Dataset, namely OULP-C1V2. This dataset comprises two main subsets A and B. A is a set of two sequences per subject and B is a set of one sequences per subject. In addition, each of the main subsets is further divided into 5 subsets based on the observation angles, 55°, 65°, 75°, 85° including all four views. The dataset consists of over 4,000 persons walking on the ground surrounded by 2 cameras at 30 fps, 640 × 480 pixels. But currently, only a form of size-normalized silhouette sequences is distributed, which are collected into two folders, Seq00 and Seq01. There are 434,731 silhouette images in Seq00 and 440,558 silhouette images in Seq01, respectively.

In this experiment, we constructed four base HMM learners and train them using gait data from four views independently. Besides, similar as the other two experiments, we designed three test cases to study the recognition rates of different methods as shown in Table 5. In all of the test cases, the split samples were randomly repeated 1000 times. The mean and standard deviation are presented in Table 6.

The CCRs based on all of the test cases are compared with the existing state-of-the-art methods in Table 6. It shows that the proposed MRGR-FF method performs well in all cases, especially in Test Case 3. This can be attributed to the well-designed static-gait features which excavate the essential characteristics of human gait. What is more significant is that the CVGR-EL method achieves 97.1% average recognition rate, much better than the second best method.

Furthermore, Fig. 8 shows CMC curves for the average recognition rate of the three test cases. Among the seven curves, the proposed CVGR-EL method shows better recognition rate compared to those existing methods. The principal reasons are the well-designed static-gait feature and the fusion of several HMM learners.

5 Conclusion

In this paper, we have investigated video-based gait recognition, which can be utilized extensively in surveillance and biometric applications. Firstly, based on area average distance, we proposed a novel method of static-feature extraction for gait recognition. Then, a view-dependent strategy is used to train each HMM gait model. Finally, we used ensemble learning for learner combination and achieved the goal of gait recognition. The comparative experiments are conducted on the well-known CASIA gait database and OU-ISIR gait database. Furthermore, we use several state-of-the-art methods, including two new methods that are specially designed to deal with the cross-view gait recognition problem. The experimental results indicate that our proposed methods are able to generate a high recognition rate for gait recognition compared with those existing approaches.

As one limitation of this work, our algorithm has only been tested on three gait datasets of indoor scenes. The gait recognition problem of outdoor scenes in more complex backgrounds needs further verification. In subsequent studies, we will study how to employ the proposed algorithms on other gait datasets and practical applications. In addition, the proposed method in this paper only uses traditional machine learning techniques and does not combine with deep learning technology. How to use the powerful feature extraction capacity of deep learning technology to further advance the recognition rate of gait recognition is another research project of our future work.

Acknowledgements We thank reviewers for their constructive suggestions and valuable comments.

Compliance with ethical standards

Conflict of interest We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled “Cross-View Gait Recognition Through Ensemble Learning”.

References

1. Sarkar S, Phillips P, Liu Z (2005) The humanID gait challenge problem: data sets, performance, and analysis. IEEE Trans Pattern Anal Mach Intell 27(02):162–177
2. Wu Z, Huang Y, Wang L, Wang X, Tan T (2017) A comprehensive study on cross-view gait based human identification with deep CNNs. IEEE Trans Pattern Anal Mach Intell 39(02):209–226
3. Zhang J, Pu J, Chen C, Fleischer R (2010) Low-resolution gait recognition. IEEE Trans Syst Man Cybern Part B (Cybernetics) 40(4):986–996
4. Guan Y, Li C, Roli F (2015) On reducing the effect of covariate factors in gait recognition: a classifier ensemble method. IEEE Trans Pattern Anal Mach Intell 37(07):1521–1529
5. Yu S, Tan D, Tan T (2006) A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In: International conference on pattern recognition, pp 441–444
6. Iwama H, Okumura M, Makihara Y, Yagi Y (2012) The OU-ISIR gait database comprising the large population dataset and performance evaluation of gait recognition. IEEE Trans Inf Forensics Secur 7(5):1511–1521
7. Moustakas Konstantinos, Tzovaras Dimitrios, Stavropoulos Georgios (2010) Gait recognition using geometric features and soft biometrics. IEEE Signal Process Lett 17(04):367–370
8. Huang X, Boulgouris N (2012) Gait recognition with shifted energy image and structural feature extraction. IEEE Trans Image Process 21(04):2256–2268
9. Kusakunniran W, Wu Q, Zhang J, Li H (2012) Gait recognition across various walking speeds using higher order shape configuration based on a differential composition model. IEEE Trans Syst Man Cybern Part B (Cybernetics) 42(6):1654–1668
10. Rida I, Jiang X, Marcialis GL (2016) Human body part selection by group lasso of motion for model-free gait recognition. IEEE Signal Process Lett 23(01):154–159
11. Theekhanont P, Kurutach W, Miguet S (2012) Gait recognition using GEI and pattern trace transform. In: International Symposium on Information Technologies in Medicine and Education, pp 936–940
12. Wang X, Wang J, Yan K (2018) Gait recognition based on Gabor wavelets and (2D)2PCA. Multimed Tools Appl 77(10):12545–12561
13. Zhang J, Pu J, Chen C, Fleischer R (2010) Low-resolution gait recognition. IEEE Trans Syst Man Cybern Part B Cybern 40(04):986–997
14. Lai Z, Xu Y, Jin Z, Zhang D (2014) Human gait recognition via sparse discriminant projection learning. IEEE Trans Circuits Syst Video Technol 24(10):1651–1662
15. Muramatsu D, Makihara Y, Yagi Y (2015) Cross-view gait recognition by fusion of multiple transformation consistency measures. IET Biom 4(2):62–73
16. Muramatsu D, Shiraishi A, Makihara Y, Uddin MZ, Yagi Y (2015) Gait-based person recognition using arbitrary view transformation model. IEEE Trans Image Process 24(1):140–154
17. Choudhury Sruti Das, Tjahjadi Tardi (2015) Robust view-invariant multiscale gait recognition. Pattern Recognit 48(03):798–811
18. Hu M, Wang Y, Zhang Z, Zhang D, Little JJ (2013) Incremental learning for video-based gait recognition with LBP flow. IEEE Trans Cybern 43(1):77–89
19. Connie T, Goh M, Teoh A (2017) A grassmannian approach to address view change problem in gait recognition. IEEE Trans Cybern 47(06):1395–1408
20. DaigoMuramatsu Makihara, Yagi Yasushi (2016) View transformation model incorporating quality measures for cross-view gait recognition. IEEE Trans Cybern 47(07):1602–1615
21. Zhang C, Liu W, Ma H, Fu H (2016) Siamese neural network based gait recognition for human identification. In: IEEE international conference on acoustics, speech and signal processing, pp 2832–2836
22. Boulgouris N, Huang X (2013) Gait recognition using HMMs and dual discriminative observations for sub-dynamics analysis. IEEE Trans Image Process 22(09):3636–3647
23. Islam M, Islam M, Hossain M, Ferworn A, Molla M (2017) Subband entropy-based features for clothing invariant human gait recognition. Adv Robot 31(10):519–530
24. Aggarwal H, Vishwakarma D (2017) Covariate conscious approach for gait recognition based upon zernike moment invariants. IEEE Trans Cogn Dev Syst 1(99):1–1
25. Wang Liang, Ning Huazhong, Tan Tieniu, Weiming Hu (2004) Fusion of static and dynamic body biometrics for gait recognition. IEEE Trans Circuits Syst Video Technol 14(2):149–158
26. Cuntoor N, Kale A, Chellappa R (2003) Combining multiple evidences for gait recognition. In: IEEE international conference on acoustics, speech, and signal processing, vol 3, p III–33
27. Veres GV, Nixon MS, Middleton L, Carter JN (2005) Fusion of dynamic and static features for gait recognition over time. In: The 7th international conference on information fusion, vol 2, pp 7–16

28. Wang Liang, Tan Tieniu, Ning Huazhong, Weiming Hu (2003) Silhouette analysis-based gait recognition for human identification. IEEE Trans Pattern Anal Mach Intell 25(12):1505–1518
29. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(5):436–445
30. Zhang H, Ji Y, Huang W, Liu L (2018) Sitcom-star-based clothing retrieval for video advertising: a deep learning framework. Neural Comput Appl 1(1):1–20
31. Takemura N, Makihara Y, Muramatsu D, Echigo T, Yagi Y (2018) On input/output architectures for convolutional neural network-based cross-view gait recognition. IEEE Trans Circuits Syst Video Technol 28(1):316–322
32. Kuncheva LI, Bezdek JC, Duin RPW (2001) Decision templates for multiple classifier fusion: an experimental comparison. Pattern Recognit 34(2):299–314
33. Wang Xuan, Wu Qingxiang, Lin Xiaojin, Zhuo Zhigiang, Huang Liuping (2016) Pedestrian identification based on fusion of multiple features and multiple classifiers. Neurocomputing 188(SI):151–159
34. Nguyen Tien Thanh, Nguyen Thi Thu Thuy, Pham Xuan Cuong, Liew Alan Wee-Chung (2016) A novel combining classifier method based on variational inference. Pattern Recognit 49:198–212
35. Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. R Stat Soc 39(1):1–22
36. Epaillard Elise, Bouguila Nizar (2016) Proportional data modeling with hidden Markov models based on generalized Dirichlet and Beta-Liouville mixtures applied to anomaly detection in public areas. Pattern Recognit 55:125–136
37. Zhang Q, Xu S (2009) Gait-based recognition of human using an embedded hidden Markov models. In: International conference on information engineering and computer science, pp 1–4
38. Shakhnarovich G, Darrell T (2002) On probabilistic combination of face and gait cues for identification. In: Proceedings of fifth IEEE international conference on automatic face gesture recognition, pp 176–181
39. Zhang H, Liu G, Chow TWS, Liu W (2011) Textual and visual content-based anti-phishing: a bayesian approach. IEEE Trans Neural Netw 22(10):1532–1546
40. McLachlan G, Krishnan T (2008) The EM algorithm and extensions. Wiley, New York
41. Zheng S, Zhang J, Huang K, He R, Tan T (2011) Robust view transformation model for gait recognition. In: The 18th IEEE international conference on image processing, pp 2073–2076