Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

A partition approach for robust gait recognition based on gait template fusion*

Kejun WANG, Liangliang LIU‡, Xinnan DING, Kaiqiang YU, Gang HU
College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin 150001, China
E-mail: heukejun@126.com; liuliangliang@hrbeu.edu.cn; dingxinnan@hrbeu.edu.cn; yukaiqiang@hrbeu.edu.cn; hugang@hrbeu.edu.cn
Received July 25, 2020; Revision accepted Nov. 11, 2020; Crosschecked Feb. 3, 2021; Published online Apr. 24, 2021

Abstract: Gait recognition has significant potential for remote human identification, but it is easily influenced by identity-unrelated factors such as clothing, carrying conditions, and view angles. Many gait templates have been presented that can effectively represent gait features. Each gait template has its advantages and can represent different prominent information. In this paper, gait template fusion is proposed to improve the classical representative gait template (such as a gait energy image) which represents incomplete information that is sensitive to changes in contour. We also present a partition method to reflect the different gait habits of different body parts of each pedestrian. The fused template is cropped into three parts (head, trunk, and leg regions) depending on the human body, and the three parts are then sent into the convolutional neural network to learn merged features. We present an extensive empirical evaluation of the CASIA-B dataset and compare the proposed method with existing ones. The results show good accuracy and robustness of the proposed method for gait recognition.

Key words: Gait recognition; Partition algorithms; Gait templates; Gait analysis; Gait energy image; Deep convolutional neural networks; Biometrics recognition; Pattern recognition
https://doi.org/10.1631/FITEE.2000377 **CLC number:** TP391.4

1 Introduction

Vision-based gait recognition is an approach that identifies humans by their walking habits, or gait, which is difficult to imitate. Recognition of gait, as one of the biometric features, has been considered a potential means of long-distance identification (Phillips, 2002; Rida et al., 2019). Unlike traditional biometrics, such as faces and fingerprints, recognition of gait needs no contact or explicit cooperation by subjects. Considering these strengths, gait recognition has become an urgent demand in the field of safety monitoring and intelligent surveillance (Iwama et al., 2013). Gait recognition has gradually attracted wide attention from researchers in the computer vision field.

The gait recognition process is more complicated than those of other biometric techniques because the input is a gait video sequence rather than a picture. In gait recognition, it is not easy to learn the features directly from the video. Most methods construct an image as a template used to represent a sequence of silhouettes, an approach for preliminary feature integration (Lv et al., 2015). There are other methods of gait recognition that learn the features directly from the video as an image set consisting of every silhouette frame (Wu et al., 2015; Chao et al., 2019). However, a gait template (such as a gait energy image (GEI)) and a set of images may be affected by contour changes or loss of temporal information concerning gait sequences (Han and Bhanu, 2006). This paper overcomes the problems mentioned above by presenting a gait template fusion method that is

‡ Corresponding author
* Project supported by the National Natural Science Foundation of China (No. 61573114)
ORCID: Liangliang LIU, https://orcid.org/0000-0002-7262-7502

based on multi-channel mapping and a novel temporal method called the fusion energy image (FEI). This proposed method has all the advantages of the existing templates, including effectively retaining the representative gait information.

There are many challenges in the practical application of gait recognition due to factors such as how pedestrian silhouettes are influenced by changes in dress, carrying conditions, and the direction of the walk (Sarkar et al., 2005). Various approaches have been proposed for this problem, but all can be divided into two typical categories, i.e., projection of the gait feature to subspace and convolutional neural networks (CNNs). Typical methods of projection are canonical correlation analysis (CCA) and coupled metric learning (CML), which can normalize gait features under different conditions into other spaces to measure their similarity (Bashir et al., 2010a, 2010b; Wang KJ et al., 2014). However, the performance of projection-based methods is affected by factors that are unrelated to identity. With the development of deep learning and the advantage of feature learning, CNNs have achieved significant improvement when applied in gait recognition (Shiraga et al., 2016; Li C et al., 2017). Generative adversarial networks (GANs) are also employed in gait recognition for normalization (Goodfellow et al., 2014), through elaborate networks that stimulate the projection to implement multiple mappings by one model (Yu et al., 2017a, 2017b; He et al., 2019). As a result of these gaps, we present a partition approach to reduce the impact of covariates, and design the CNN structure based on the merging of features for the application of gait recognition.

In summary, we make the following contributions in comparison with the state-of-the-art methods:

1. A gait template fusion method based on multi-channel mapping and a novel temporal approach called FEI is proposed to effectively represent more original gait information. It can combine the advantages of several existing gait representative templates simply and efficiently while maintaining a low computational cost.

2. We propose a partition approach based on the structure of the gait contour. The FEI is separated into three parts: head, trunk, and leg. The features of each part of the body are learned from each block, and the low-dimensional features of the GEI are obtained after fusion.

3. A CNN-based fusion method is presented for gait feature extraction and recognition. Bottom layers with three routes in parallel are used to learn the gait features, and then identity recognition is performed through the fusion layer.

4. We conduct an extensive evaluation under different clothing, carrying conditions, and view angles. Gait recognition accuracy is improved by the proposed method on the CASIA-B dataset.

2 Related works

Most gait recognition methods involve various steps: video information integration, feature extraction, and similarity metric classification (Wang C et al., 2012). The existing works are presented based on different methods of video information integration and cross-domain feature learning.

2.1 Gait representative template

Due to the difficulty in directly obtaining information from videos, it is essential to convert a sequence of images into a single representative template in gait recognition. Han and Bhanu (2006) presented GEI which is the most classic template. The GEI was calculated by superimposing the average of periodic normalized silhouettes while losing temporal information. Zhang EH et al. (2010) used an active energy image (AEI), consisting of the frame difference image, which is a good response to the temporal features of the moving part. Although it is not sensitive to image quality, it has less contour information. Bashir et al. (2010a) proposed a gait entropy image (GEnI) by calculating the Shannon entropy of each pixel of a GEI and the density of different positions of the profile by Shannon entropy. It effectively solves the problem of sensitivity due to the change of contour of the pedestrian under different clothing and carrying conditions. Based on the idea of GEI, Wang C et al. (2012) applied a multi-channel mapping function to encode temporal information to project the silhouettes onto the RGB space, and named this template the Chrono-gait image, which is still sensitive to changes in clothing and carrying conditions.

2.2 Cross-domain gait features learning method

Cross-domain gait recognition has always been a big challenge due to the huge change of contour caused by the view angle, clothing, and carrying conditions. Various approaches have been proposed to solve this problem. They can be divided into two classes, i.e., projection-based and CNN-based.

Projection-based methods are always based on the distance metric of gait features. Bashir et al. (2010a) applied CCA to a model with different view angles. Xing et al. (2016) developed a complete canonical correlation analysis method (C3A) to deal with the unsteady and incomplete singular matrix problem of generalized feature analysis in the CCA framework of two high-dimensional datasets. It has also been reported that a view transformation model is a typical method, where gait features are projected from one view domain to another. The gait matrix is decomposed into a view-independent matrix and an object-independent matrix using singular value decomposition (Makihara et al., 2006; Kusakunniran et al., 2012; Muramatsu et al., 2016). CML is an effective gait recognition approach for detecting the essential structure of tensor manifolds by preserving local information to encode intra-class compactness and inter-class separability with local relationships (Wang KJ et al., 2014; Ben et al., 2019a, 2019b). Ben et al. (2020) presented a coupled bilinear discriminant projection to reduce the gap caused by view difference based on a coupled distance metric without losing the GEIs' spatial information.

Recently, CNNs have become the common workhorse for feature learning from images. Constructing a CNN is helpful in extracting gait features from a sequence of silhouette sets (Wu et al., 2015; Li C et al., 2017; Chao et al., 2019). They can also be learned through cross-domain features using gait templates. Wu et al. (2017) designed a CNN structure for gait recognition based on a GEI with good cross-view performance. Yu et al. (2017b) employed a deep model using stacked progressive auto-encoders (SPAE) to convert the gait features from one view to another to reduce the influence of view changes. Song et al. (2019) deployed joint learning to combine silhouette extraction and gait recognition in an end-to-end framework.

In addition, GANs restore the probe and the corresponding gallery data standard condition to learn invariant gait features by combining projection methods and CNNs. Yu et al. (2017b) employed a GAN as a regressor to generate invariant gait images while preserving human identification, converting arbitrary gait images to the side view with normal clothing but without carrying objects. Zhang P et al. (2019) proposed the variation-normalizing GAN, deploying a coarse-to-fine method, where initial coarse images were generated by normalizing the view and then refined by injecting identity-related information. Li X et al. (2020) presented an alpha-blending GAN to remove carried objects to obtain identity-preserved alpha-blended gait templates.

3 Method

3.1 Overview

In this study, we present a novel fusion-based gait recognition method, taking the advantage of more gait information and solving the problem posed by silhouettes in different view angles or dresses. The application of the fusion method is reflected in two ways. First, by fusing existing gait representative templates, the original information is fully used while maintaining a low computational cost. Second, the features of each partitioned sub-FEI are learned by a CNN, and the final gait features are obtained by a feature fusion method for classification of the nearest neighbor (NN).

Fig. 1 shows the systematic workflow of the proposed method; the algorithm is roughly understood as three main parts, i.e., gait representative template fusion, partition, and feature fusion based on convolutional networks. The following subsections present these three parts in detail.

3.2 Gait representative template fusion method based on multi-channel mapping

The existing single gait representative template effectively integrates the gait video information but cannot avoid losing features. For example, GEI loses gait timing information and AEI loses contour features. However, each template has a unique advantage, such as insensitivity to changes in dress with AEI. Because most templates are gray-level images, it is reasonable to create a colorful new fusion template

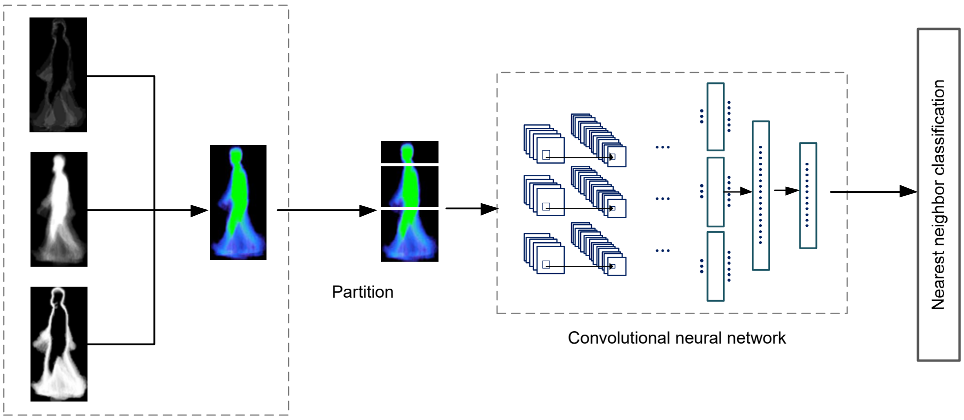

Fig. 1 Process flow of the proposed method

that will represent more gait information with low calculation costs.

The multi-channel mapping method is used to complete the fusion of multiple templates. Gait templates are marked with color and mapped to the RGB space in the red, green, and blue channels. For M fused templates, the weights of the m^{th} template in the RGB space are

$$i_m = (m-1)/(M-1), \tag{1}$$

$$W_m^{\mathrm{R}} = \begin{cases} 1-2i_m, & 0 \le i_m \le 1/2, \\ 0, & 1/2 < i_m < 1, \end{cases} \tag{2}$$

$$W_m^{\mathrm{G}} = \begin{cases} 0, & 0 \le i_m \le 1/2, \\ 2i_m - 1, & 1/2 < i_m < 1, \end{cases} \tag{3}$$

$$W_m^{\mathrm{B}} = \begin{cases} 2i_m, & 0 \le i_m \le 1/2, \\ 2-2i_m, & 1/2 < i_m < 1, \end{cases} \tag{4}$$

where i_m is the normalized number of the m^{th} template, and $W_m^{\mathrm{R}}, W_m^{\mathrm{G}}$, and W_m^{B} are the weights mapped to the RGB space. So, $\mathbf{FEI}(x, y)$ is defined as follows:

$$\mathbf{FEI}(x,y) = \begin{pmatrix} T_1(x,y)W_1^{\mathrm{R}} + T_2(x,y)W_2^{\mathrm{R}} + \cdots + T_m(x,y)W_m^{\mathrm{R}} \\ T_1(x,y)W_1^{\mathrm{G}} + T_2(x,y)W_2^{\mathrm{G}} + \cdots + T_m(x,y)W_m^{\mathrm{G}} \\ T_1(x,y)W_1^{\mathrm{B}} + T_2(x,y)W_2^{\mathrm{B}} + \cdots + T_m(x,y)W_m^{\mathrm{B}} \end{pmatrix}, \tag{5}$$

where $T_m(x, y)$ is the image of the m^{th} template. In this study, we choose GEI, AEI, and GEnI to compose the FEI. GEI has an efficient gait characterization template, but has no connection between adjacent frames. AEI can extract the active regions by calculating the frame difference. Note that GEnI is not sensitive to changes in static information about carrying conditions or clothing. The FEI composed of these three templates contains more gait information and the respective shortcomings are offset by each other.

As the most widely used gait template, GEI has an effective representation of gait sequences and high computational efficiency, i.e., the energy of a superposition of a period of normalized silhouettes (Han and Bhanu, 2006). The GEI is calculated as

$$E_{\mathrm{GEI}}(x,y) = \frac{1}{N}\sum_{n=1}^{N} B(x,y,n), \tag{6}$$

where N is the number of frames in a gait cycle and $B(x, y, n)$ is the binary and normalized silhouette of the n^{th} frame. The concentration of the pixel, i.e., energy, represents the frequency of human motion at this pixel, as shown in the third row of Fig. 2. GEI is a gray-scale image that represents the dynamic and static (shape) information of a gait sequence, although it is not sensitive to occasional errors in a single frame.

GEI is an effective gait template, but does not change the motion between adjacent frames. To solve this problem, Zhang EH et al. (2010) proposed AEI, which can extract the active regions by calculating the frame difference. AEI is defined as follows:

$$E_{\mathrm{AEI}}(x,y) = \frac{1}{N}\sum_{n=1}^{N} D_n(x,y,n), \tag{7}$$

$$D_n(x,y,n)=\left|B(x,y,n+1)-B(x,y,n)\right|, \tag{8}$$

where $D_n(x, y, n)$ is the difference between the next frame and the current frame of the silhouettes. There are more temporal characteristics in AEI than in GEI from the comparison of rows 1 and 3 in Fig. 2.

As shown in Fig. 2, AEI, GEI, and GEnI are displayed in the three channels of the FEI accordingly. Furthermore, the influence of carrying conditions or clothing is reduced, but AEI ignores static information by using only the dynamic part of the silhouettes.

To increase dynamic information, dynamic and static regions are distinguished from GEI by measuring the Shannon entropy of each pixel position in the GEI (Bashir et al., 2010a). By strengthening the dynamic region in this way, the GEnI can be obtained:

$$\begin{aligned}E_{\mathrm{GEnI}}(x,y)=&-E_{\mathrm{GEI}}(x,y)\log_2 E_{\mathrm{GEI}}(x,y)\\&-(1-E_{\mathrm{GEI}}(x,y))\log_2(1-E_{\mathrm{GEI}}(x,y)).\end{aligned} \tag{9}$$

Due to the uncertainty of the dynamic region, the intensity of the GEnI is higher in the dynamic region than in the static region, as shown in the second row of Fig. 2.

FEI is calculated according to Eq. (5) as shown in the last row in Fig. 2. At this point, M is three. It is easy to calculate the weights of the three channels of AEI, GEI, and GEnI, and FEI can be described as representing all information about the three templates. The GEI has a satisfactory extraction of contour features, but the dynamic information is insufficiently expressed and the timing information is lost. Therefore, there is no disadvantage in the FEI because AEI and GEnI make up for the deficiency of GEI.

$$\mathbf{FEI}(x,y)=\begin{pmatrix}E_{\mathrm{AEI}}(x,y)\\E_{\mathrm{GEI}}(x,y)\\E_{\mathrm{GEnI}}(x,y)\end{pmatrix}. \tag{10}$$

3.3 A robust partition approach for covariate factors

In general, gait information is based on the sequence of the entire silhouette. However, the covariate factors, such as clothes and bags, make the appearance of the silhouette change so greatly that it is difficult to extract gait information effectively. As illustrated in Fig. 2, the three FEIs belonging to the same pedestrian show a great difference in silhouette, and this difference has a great impact on the accuracy of vision-based gait recognition.

As shown in Fig. 3, the changing of silhouettes

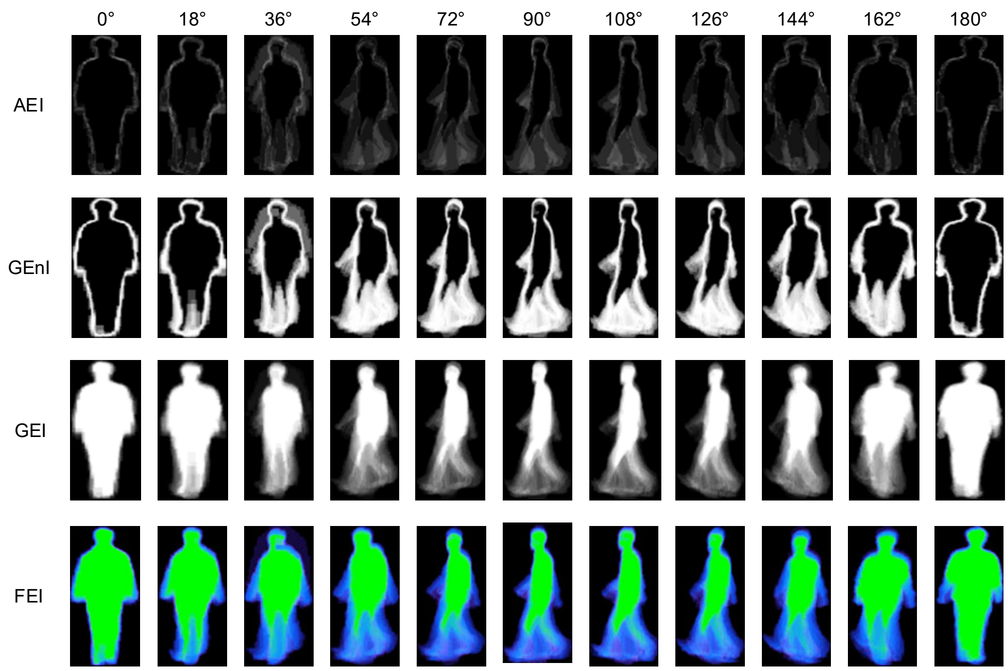

Fig. 2 AEIs, GEnIs, GEIs, and FEIs at different view angles

caused by the clothing or backpacks is concentrated mainly on the human body's torso, and less on the head and legs. In the traditional method, based on the whole image, each human body part contributes to the recognition results, which is clearly inappropriate. Also, some existing works demonstrate that learning from local images is conducive to enhancing feature extraction (Hossain et al., 2010; Wang YY et al., 2019; Fan et al., 2020).

Therefore, we present a solution to the covariate factors based on partition. FEI is cropped into the head, trunk, and leg regions as shown in Fig. 3. Due to normalization in the FEI generation process, it is easy to apply this method to all the subjects. In this way, attention is not solely focused on the whole silhouette, but also dispersed to each sub-FEI to reduce the effect of covariate factors. The size of the FEI is 64×128×3, and the sizes of the three sub-FEIs are 64×32×3, 64×48×3, and 64×48×3.

3.4 Feature-level information fusion based deep convolutional networks

Feature-level fusion is designed to extract features to obtain information. The purpose is to increase the number of features in the FEI. The CNN can learn local features and silhouettes together, and hence it can use the information of given features and improve recognition performance. Three sub-FEIs are sent to the CNN to learn the gait feature, and the fusion layer combines multiple features to recognize it.

Table 1 shows the structure of the network in detail. The output shape of each layer is given after the structure. Bottom layers with three routes in parallel are used to learn the gait features, respectively. The middle layer is the fusion layer with top layers for classification. Conv5 has 5×5 kernels, and Conv3 has 3×3 kernels. Similarly, maxpool2 means the maximum pool layer with 2×2 kernels. Moreover, all of the activation functions are ReLU.

The softmax loss is used for network training, classifying each gait pose into the corresponding subject and enlarging the inter-class distance effectively. It is defined as

$$L=-\sum_{i=1}^{m}\log\left(\frac{\mathrm{e}^{\boldsymbol{W}_{y_i}^{\mathrm{T}}\boldsymbol{x}_i+\boldsymbol{b}_{y_i}}}{\sum_{j=1}^{n}\mathrm{e}^{\boldsymbol{W}_j^{\mathrm{T}}\boldsymbol{x}_i+\boldsymbol{b}_j}}\right), \tag{11}$$

where $\boldsymbol{x}_i\in\mathbb{R}^d$ is the i^{th} feature that belongs to the y_i^{th} class, d denotes the feature dimension, $\boldsymbol{W}_{y_i}, \boldsymbol{W}_j\in\mathbb{R}^{d\times n}$ denote the last connected layers, and $\boldsymbol{b}_{y_i}, \boldsymbol{b}_j\in\mathbb{R}^n$ denote bias terms.

The GEI is divided into three blocks according to body parts, and they are sent to the convolutional network to learn gait features accordingly. Not only can the CNN learn global gait silhouettes, but also local body part information is extracted.

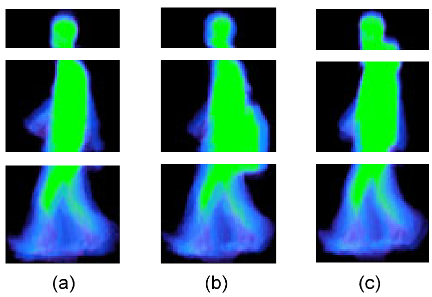

Fig. 3 Partitioned FEI when the person is in the normal condition (a), carrying a bag (b), and wearing a coat (c)

Table 1 Implementation details of the network

Layer (head)	Layer (trunk)	Layer (leg)
Conv5-ReLU-Batch_norm (60, 28, 16)	Conv5-ReLU-Batch_norm (60, 44, 16)	Conv5-ReLU-Batch_norm (60, 44, 16)
	Conv3-ReLU (58, 42, 32)	Conv3-ReLU (58, 42, 32)
Maxpool2 (30, 14, 16)	Maxpool2 (29, 21, 32)	Maxpool2 (29, 21, 32)
Conv3-ReLU-Batch_norm (28, 12, 32)	Conv3-ReLU-Batch_norm (27, 19, 64)	Conv3-ReLU-Batch_norm (27, 19, 64)
	Conv3-ReLU (25, 17, 64)	Conv3-ReLU (25, 17, 64)
Maxpool2 (14, 6, 32)	Maxpool2 (14, 8, 64)	Maxpool2 (14, 8, 64)
FC-ReLU (512)	FC-ReLU (1024)	FC-ReLU (1024)
Concat (2560)	Concat (2560)	Concat (2560)
FC-ReLU (1024)	FC-ReLU (1024)	FC-ReLU (1024)
FC-ReLU (512)	FC-ReLU (512)	FC-ReLU (512)
FC-ReLU (62)	FC-ReLU (62)	FC-ReLU (62)

4 Experiments and analysis

An empirical evaluation was provided on the CASIA-B gait dataset. There were 124 subjects and 11 view angles (0°, 18°, …, 180°), and 10 sequences per subject for each view angle in the CASIA-B gait dataset (Yu et al., 2006). Six of the 10 sequences were taken under normal walking conditions (four of them were selected as galleries with the other two as probes (NM)); two sequences when the subjects wear their coats were used as probes (CL); the remaining two with bags were kept as probes (BG). We evaluated our method by comparing it to alternative approaches in terms of different clothing, carrying conditions, and view angles.

The CASIA-B database provides segmented binary gait contours, but some image preprocessing steps are still needed before carrying out the experiment. We need to normalize all the frames so the contours are of equal size to avoid the influence of different distances between the human and camera to generate a gait representative template. In this process, the top and bottom pixels of the silhouettes should be located to establish the pedestrian area, and then their centers of gravity were calculated. After calculating the center of gravity, the height of the image, and the aspect ratio (1/2), the frames were cropped and resized to 64×128.

The first 62 subjects of the 124 were put in the training set, and the four normal walking sequences of the remaining subjects were kept as galleries in the test set. The networks were trained using stochastic gradient descent with a learning rate of 0.001. The batch size was set to 128, and there were about 30 000 iterations.

4.1 Impact of fusing gait representative templates

We designed a simple experiment to prove the validity of the template fusion algorithm. From Table 2, the experiment is provided with AEI, GEI, GEnI, and FEI (consisting of these three) as gait representative templates. The NN classifier was used for identification to establish accuracy, and the view angle of both the gallery and probe was 90°.

The accuracy of recognition with FEI is the best in all three conditions with more than 2% improvement over GEI (Table 2). Moreover, it can be seen that more information is lost in AEI, which is less affected by covariate factors as compared to the experimental results of the same template under different conditions. There are more features and a slight improvement in GEnI than in GEI when clothing worn or carrying an object. FEI effectively synthesizes the information of three templates and achieves better recognition results. None of the recognition rates are high in the experiment. This may be due to the weak robustness caused by the simple feature extraction approach and classifier.

Table 2 Comparison of different gait representative templates by accuracy

Probe	Accuracy (%)			
	AEI	GEI	GEnI	FEI
NM	88.7	93.5	94.8	96.8
CL	35.2	32.7	34.3	35.4
BG	46.3	44.5	45.8	47.0

Gallery: NM 1–4, 90°; probe: NM 5–6, CL, BG

4.2 Impact of CNN-based partition and fusion

In this subsection, we evaluate the effectiveness of partition and fusion based on the CNN. Fig. 4 compares the recognition accuracies of the two approaches, with or without partition. As a result of the partitioned FEI with a dedicated network structure, the CNN of the FEI structure is similar to it without the bottom layers of three parallel routes and the intermediate fusion layer. The setting of gallery and probe is consistent with the experiment in Section 4.1.

The effectiveness of the proposed partition method is illustrated in Fig. 4. Both of them achieved more than 98% accuracy in normal walking because the CNN has the advantages of good feature learning and generalization performance. The partition and fusion method's performance is slightly better when there are changes in the clothing or carrying conditions, showing that the proposed method increases the robustness of covariate factors.

The partition methods and the number of partitions influence our results. Table 3 shows a comparison with different numbers of partitions (partitions are based on the human body structure).

We observed that partition approaches perform better in the clothing worn and bag-carrying conditions, which means that this method is effective in dealing with covariate factors. Additionally, three-sub-FEI is the best method in the clothing condition, whereas two-sub-FEI is the best method in the bag-carrying condition. This may be because these two methods both split the trunk and legs, and are more successful at extracting local features. When the number of sub-FEIs is four, the result becomes worse due to dense separation, which is detrimental to global learning.

4.3 Comparison with other methods

We evaluate the accuracy of the proposed FEI+CNN and FEI+partition+CNN methods, and the effect of view and covariate variations. The results of the comparison are shown in Figs. 5–7, with the state-of-the-art cross-condition gait recognition methods including principal component analysis and linear discriminant analysis (GEI+PCA+LDA), CCA (Bashir et al., 2010b), SPAE (Yu et al., 2017b), and VN-GAN (Zhang P et al., 2019). Cross-condition refers to gait recognition where the conditions (such as view angle, clothing worn, and carrying an object) of galleries and probes are not the same. The GEI+PCA+LDA method is easy to understand, where PCA is applied for feature extraction with LDA for identity recognition. CCA is applied to model gait sequences from different conditions with the similarity measure of correlation strength (Bashir et al., 2010a). SPAE is based on multi-layer auto-encoders that extract the invariant gait feature of different conditions (Yu et al., 2017b). VN-GAN employs a divide-and-conquer strategy where coarse, view-normalized gaits are generated first, and then identity information is implanted (Zhang P et al., 2019).

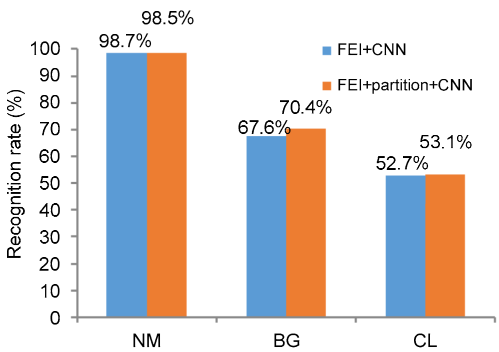

Fig. 4 Comparison of partition and fusion by accuracy (References to color refer to the online version of this figure)
Gallery: NM 1–4, 90°; probe: NM 5–6, CL, BG

Fig. 5 shows the results in comparison with the existing methods in normal walking, where the view angles of the gallery are 0°–180°. A significant change of view angle is a big challenge for gait recognition. The proposed methods are significantly better than almost all other algorithms. This can be seen in the FEI-based approach, which performs better than the partitioned FEI in most view angles. This is because the FEI-based approach may focus more on the overall contour information, which is more useful in the normal walking condition of the gallery and probe. Notably, our results do not support the general understanding that the accuracy at 90° is not the best. This may be due to gait recognition based on silhouettes. It is unable to discriminate the left- or right-hand side of the body structure in silhouettes, thus causing confusion, such as the left or right leg as shown in Fig. 1, while more stereoscopic information is included in the better performing 36° and 144°.

Experimental settings are unchanged except for the condition of probes in coats or with a bag. Figs. 6 and 7 show the performance comparisons of the existing approaches under more complex conditions. The superiority of our method is embodied in the condition of complex changes. Moreover, the method with partitioned FEI achieves higher accuracy in almost all view angles than the FEI-based approach, revealing that the partition and fusion methods are robust to covariate factors. It can also be seen that the accuracy is always higher in the condition of carrying a bag than wearing a coat. The reason may be that the

Table 3 Comparison of accuracy at different numbers of partitions

Probe	Accuracy (%)			
	Partition number=1	Partition number=2	Partition number=3	Partition number=4
NM	98.7	98.5	98.5	98.0
CL	67.6	70.1	70.4	68.2
BG	52.7	53.3	53.1	53.1

Gallery: NM 1–4, 90°; probe: NM 5–6, CL, BG. Partition number=1: the whole image without partition; Partition number=2: two partition parts involve double 64×64×3 sub-FEIs; Partition number=3: three sub-FEIs are 64×32×3, 64×48×3, and 64×48×3; Partition number=4: four partition parts involve 64×32×3, 64×24×3, 64×24×3, and 64×48×3 sub-FEIs

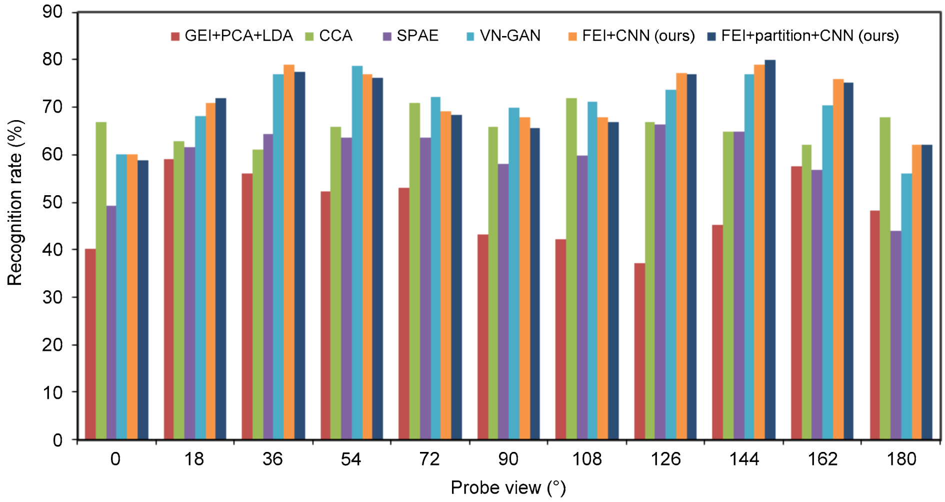

Fig. 5 Comparison of the proposed method with existing ones by accuracy under probe NM 5–6, 0°–180° (gallery: NM 1–4, 0°–180°) (References to color refer to the online version of this figure)

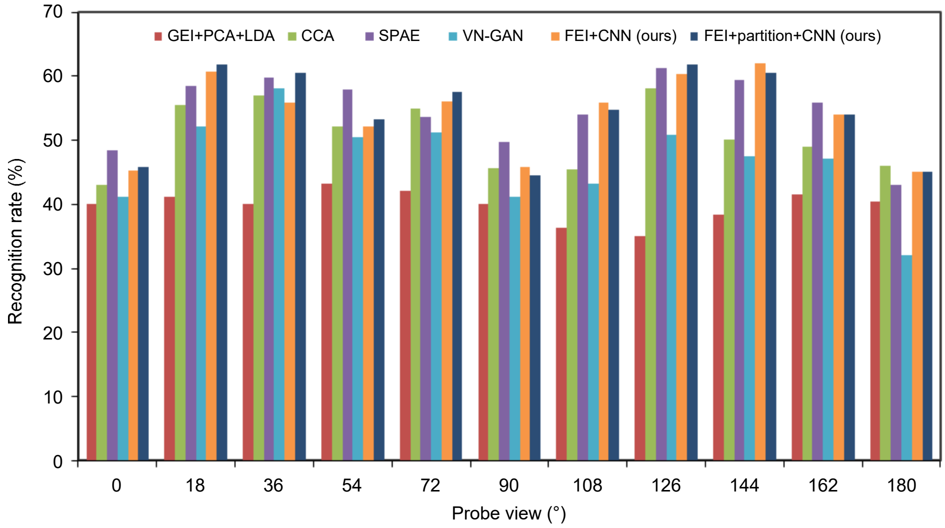

Fig. 6 Comparison of the proposed method with existing ones by accuracy under probe BG 0°–180° (gallery: NM 1–4, 0°–180°) (References to color refer to the online version of this figure)

bag only blocks out or changes part of the contour, while the effect of the clothing worn on the contour is enormous and creates more noise.

5 Conclusions

In this paper, a novel gait template fusion and partition method has been proposed for gait recognition. We have employed gait template fusion using multi-channel mapping to obtain an FEI to represent more original gait features. The FEI approach also combines the advantages of the existing templates to increase the robustness of identity-unrelated factors. Moreover, the partition and fusion method has been applied to a CNN, where the features of a partitioned FEI and fusion for identification have been extracted. This is effective in reducing the impact of changes in clothing and carrying conditions. Experimental results confirmed the proposed method's effectiveness and robustness for gait recognition, and showed especially good performance in various complex environments.

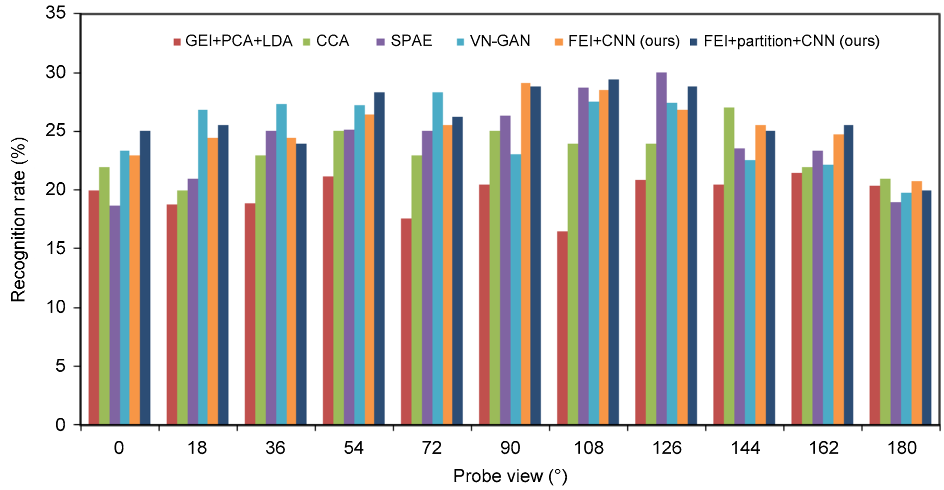

Fig. 7 Comparison of the proposed method with existing ones by accuracy under probe CL 0°–180° (gallery: NM 1–4, 0°–180°) (References to color refer to the online version of this figure)

Contributors

Kejun WANG and Liangliang LIU designed the research. Liangliang LIU, Xinnan DING, Kaiqiang YU, and Gang HU processed the data. Liangliang LIU drafted the manuscript. Kejun WANG helped organize the manuscript. Liangliang LIU and Xinnan DING revised and finalized the paper.

Compliance with ethics guidelines

Kejun WANG, Liangliang LIU, Xinnan DING, Kaiqiang YU, and Gang HU declare that they have no conflict of interest.

References

Bashir K, Xiang T, Gong SG, 2010a. Cross-view gait recognition using correlation strength. Proc British Machine Vision Conf, p.1-11. https://doi.org/10.5244/C.24.109

Bashir K, Xiang T, Gong SG, 2010b. Gait recognition without subject cooperation. *Patt Recogn Lett*, 31(13):2052-2060. https://doi.org/10.1016/j.patrec.2010.05.027

Ben XY, Gong C, Zhang P, et al., 2019a. Coupled patch alignment for matching cross-view gaits. *IEEE Trans Image Process*, 28(6):3142-3157. https://doi.org/10.1109/TIP.2019.2894362

Ben XY, Zhang P, Lai ZH, et al., 2019b. A general tensor representation framework for cross-view gait recognition. *Patt Recogn*, 90:87-98. https://doi.org/10.1016/j.patcog.2019.01.017

Ben XY, Gong C, Zhang P, et al., 2020. Coupled bilinear discriminant projection for cross-view gait recognition. *IEEE Trans Circ Syst Video Technol*, 30(3):734-747. https://doi.org/10.1109/TCSVT.2019.2893736

Chao HQ, He YW, Zhang JP, et al., 2019. GaitSet: regarding gait as a set for cross-view gait recognition. Proc AAAI Conf on Artificial Intelligence, p.8126-8133. https://doi.org/10.1609/aaai.v33i01.33018126

Fan C, Peng YJ, Cao CS, et al., 2020. GaitPart: temporal part-based model for gait recognition. Proc IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.14213-14221. https://doi.org/10.1109/CVPR42600.2020.01423

Goodfellow IJ, Pouget-Abadie J, Mirza M, et al., 2014. Generative adversarial nets. Proc 27th Int Conf on Neural Information Processing Systems, p.2672-2680.

Han J, Bhanu B, 2006. Individual recognition using gait energy image. *IEEE Trans Patt Anal Mach Intell*, 28(2):316-323. https://doi.org/10.1109/TPAMI.2006.38

He YW, Zhang JP, Shan HM, et al., 2019. Multi-task GANs for view-specific feature learning in gait recognition. *IEEE Trans Inform Forens Secur*, 14(1):102-113. https://doi.org/10.1109/TIFS.2018.2844819

Hossain A, Makihara Y, Wang JQ, et al., 2010. Clothing-invariant gait identification using part-based clothing categorization and adaptive weight control. *Patt Recogn*, 43(6):2281-2291. https://doi.org/10.1016/j.patcog.2009.12.020

Iwama H, Muramatsu D, Makihara Y, et al., 2013. Gait verification system for criminal investigation. *IPSJ Trans Comput Vis Appl*, 5:163-175. https://doi.org/10.2197/ipsjtcva.5.163

Kusakunniran W, Wu Q, Zhang J, et al., 2012. Gait recognition under various viewing angles based on correlated motion regression. *IEEE Trans Circ Syst Video Technol*, 22(6):966-980. https://doi.org/10.1109/TCSVT.2012.2186744

Li C, Min X, Sun SQ, et al., 2017. DeepGait: a learning deep convolutional representation for view-invariant gait recognition using joint Bayesian. *Appl Sci*, 7(3):210. https://doi.org/10.3390/app7030210

Li X, Makihara Y, Xu C, et al., 2020. Gait recognition invariant to carried objects using alpha blending generative adversarial networks. *Patt Recogn*, 105:107376.

https://doi.org/10.1016/j.patcog.2020.107376

Lv ZW, Xing XL, Wang KJ, et al., 2015. Class energy image analysis for video sensor-based gait recognition: a review. *Sensors*, 15(1):932-964. https://doi.org/10.3390/s150100932

Makihara Y, Sagawa R, Mukaigawa Y, et al., 2006. Gait recognition using a view transformation model in the frequency domain. Proc European Conf on Computer Vision, p.151-163. https://doi.org/10.1007/11744078_12

Muramatsu D, Makihara Y, Yagi Y, 2016. View transformation model incorporating quality measures for cross-view gait recognition. *IEEE Trans Cybern*, 46(7):1602-1615. https://doi.org/10.1109/TCYB.2015.2452577

Phillips PJ, 2002. Human identification technical challenges. Proc Int Conf on Image Processing, p.49-52. https://doi.org/10.1109/ICIP.2002.1037956

Rida I, Almaadeed N, Almaadeed S, 2019. Robust gait recognition: a comprehensive survey. *IET Biometr*, 8(1):14-28. https://doi.org/10.1049/iet-bmt.2018.5063

Sarkar S, Phillips PJ, Liu Z, et al., 2005. The humanid gait challenge problem: data sets, performance, and analysis. *IEEE Trans Patt Anal Mach Intell*, 27(2):162-177. https://doi.org/10.1109/TPAMI.2005.39

Shiraga K, Makihara Y, Muramatsu D, et al., 2016. GEINet: view-invariant gait recognition using a convolutional neural network. Proc Int Conf on Biometrics, p.1-8. https://doi.org/10.1109/ICB.2016.7550060

Song CF, Huang YZ, Huang Y, et al., 2019. GaitNet: an end-to-end network for gait based human identification. *Patt Recogn*, 96:106988. https://doi.org/10.1016/j.patcog.2019.106988

Wang C, Zhang JP, Wang L, et al., 2012. Human identification using temporal information preserving gait template. *IEEE Trans Patt Anal Mach Intell*, 34(11):2164-2176. https://doi.org/10.1109/TPAMI.2011.260

Wang KJ, Xing XL, Yan T, et al., 2014. Couple metric learning based on separable criteria with its application in cross-view gait recognition. Proc 9th Chinese Conf on Biometric Recognition, p.347-356. https://doi.org/10.1007/978-3-319-12484-1_39

Wang YY, Song CF, Huang Y, et al., 2019. Learning view invariant gait features with Two-Stream GAN. *Neurocomputing*, 339:245-254. https://doi.org/10.1016/j.neucom.2019.02.025

Wu ZF, Huang YZ, Wang L, 2015. Learning representative deep features for image set analysis. *IEEE Trans Multim*, 17(11):1960-1968. https://doi.org/10.1109/TMM.2015.2477681

Wu ZF, Huang YZ, Wang L, et al., 2017. A comprehensive study on cross-view gait based human identification with deep CNNs. *IEEE Trans Patt Anal Mach Intell*, 39(2):209-226. https://doi.org/10.1109/TPAMI.2016.2545669

Xing XL, Wang KJ, Yan T, et al., 2016. Complete canonical correlation analysis with application to multi-view gait recognition. *Patt Recogn*, 50:107-117. https://doi.org/10.1016/j.patcog.2015.08.011

Yu SQ, Tan DL, Tan TN, 2006. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. Proc 18th Int Conf on Pattern Recognition, p.441-444. https://doi.org/10.1109/ICPR.2006.67

Yu SQ, Chen HF, Reyes EBG, et al., 2017a. GaitGAN: invariant gait feature extraction using generative adversarial networks. Proc IEEE Conf on Computer Vision and Pattern Recognition Workshops, p.532-539. https://doi.org/10.1109/CVPRW.2017.80

Yu SQ, Chen HF, Wang Q, et al., 2017b. Invariant feature extraction for gait recognition using only one uniform model. *Neurocomputing*, 239:81-93. https://doi.org/10.1016/j.neucom.2017.02.006

Zhang EH, Zhao YW, Xiong W, 2010. Active energy image plus 2DLPP for gait recognition. *Signal Process*, 90(7):2295-2302. https://doi.org/10.1016/j.sigpro.2010.01.024

Zhang P, Wu Q, Xu JS, 2019. VN-GAN: identity-preserved variation normalizing GAN for gait recognition. Proc Int Joint Conf on Neural Networks, p.1-8. https://doi.org/10.1109/IJCNN.2019.8852401