DeformGait: Gait Recognition under Posture Changes using Deformation Patterns between Gait Feature Pairs

Chi Xu[1,2] Daisuke Adachi[2] Yasushi Makihara[2] Yasushi Yagi[2] Jianfeng Lu[1]
[1] Nanjing University of Science and Technology, Nanjing, China [2] Osaka University, Osaka, Japan
{xu, adachi, makihara, yagi}@am.sanken.osaka-u.ac.jp lujf@mail.njust.edu.cn

Abstract

In this paper, we propose a unified convolutional neural network (CNN) framework for robust gait recognition against posture changes (e.g., those induced by walking speed changes). In order to mitigate the posture changes, we first register an input matching pair of gait features with different postures by a deformable registration network, which estimates a deformation field to transform the input pair both into their intermediate posture. The pair of the registered features is then fed into a recognition network. Furthermore, ways of the deformation (i.e., deformation patterns) can differ between the same subject pairs (e.g., only posture deformation) and different subject pairs (e.g., not only posture deformation but also body shape deformation), which implies the deformation pattern can be another cue to distinguish the same subject pairs from the different subject pairs. We therefore introduce another recognition network whose input is the deformation pattern. Finally, the deformable registration network, and the two recognition networks for the registered features and the deformation patterns, constitute the whole framework, named DeformGait, and they are trained in an end-to-end manner by minimizing a loss function which is appropriately designed for each of verification and identification scenario. Experiments on the publicly available dataset containing the largest speed variations demonstrate that the proposed method achieves the state-of-the-art performance in both identification and verification scenarios.

1. Introduction

Gait recognition is one of behavioral biometrics which identifies a person based on how he/she walks. Gait has its unique advantages over other physiological biometrics (e.g., DNA, fingerprint, vein, and face): recognition without subject cooperation; recognition at a large distance from a camera with low-resolution images (e.g., a video captured by a CCTV in a public area). Gait recognition is therefore suitable for many applications in surveillance, forensics, and criminal investigation [3, 11, 22].

On the other hand, gait recognition performance often degrades due to various covariates, such as views [25, 18, 43], clothing [20, 21], carriages [36, 26], and walking speed [27, 7, 44], since the covariates cause large intra-subject variations on gait features, particularly on appearance-based ones (e.g., gait energy image (GEI) [9]). In addition to the above-mentioned covariates, posture change is also considered as a covariate that often occurs while walking, and makes gait recognition difficult. For example, people may sometimes look down to watch a mobile phone, or swing their arms more than usual when feeling happy [38]. Moreover, there exist posture changes caused by walking speed changes: larger stride length and more arm swing when walking faster; forward-bent posture and more knee flexion when running (see Fig. 3).

A possible solution to gain robustness against the posture changes, is to apply discriminative approaches to an image-based gait feature such as metric learning frameworks [9, 47, 28, 39, 6] or deep learning frameworks, more specifically, convolutional neural network (CNN) frameworks [42, 31, 43, 48, 34, 10, 4, 51]. In most of the metric learning frameworks, lower weights are assigned to a spatial region of the image-based gait feature where intra-subject variations affects more (e.g., a region around a torso's contour which is affected by intra-subject forward bending when running), and vice versa. The CNN frameworks usually contain max-pooling layers, which make extracted features spatially invariant more or less [34]. The spatially invariant property brought by the approaches may, however, sometimes wash out subtle yet informative inter-subject variations (e.g., torso's contour differences between normal and slightly slim subjects).

Another direction is to apply deformable registration frameworks before matching [5, 23, 46], and is considered as a more promising solution because it can directly handle the geometric aspect of the intra-subject posture changes. In particular, free form deformation (FFD) [30] is one of the most

978-1-7281-9186-7/20/$31.00

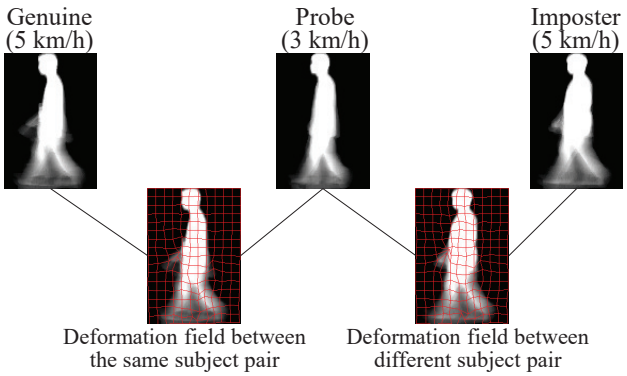

Figure 1. Deformation fields between a pair of the same subject (i.e., probe and genuine GEIs), and that of different subjects (i.e., probe and imposter GEIs) obtained by the method [46].

flexible deformation representation for a non-rigid object such as a human body, and hence it is often employed in gait analysis studies to register gait features between different views [5, 46], gait modes (i.e., walking and running) [45], and postures [23]. However, if we register the gait features too much, the aforementioned subtle yet informative inter-subject variations may be washed out (e.g., registration from a normal body shape to a slightly slim one).

By taking a closer look at the deformation patterns, we notice that there are considerable differences between intra-subject (i.e., the same subject) deformation patterns and inter-subject (i.e., different subjects) ones, as shown in Fig. 1. In case of walking speed change, the intra-subject deformation patterns are mainly derived from the posture changes of stride and arm swing, whereas inter-subject ones contain spatial displacements due to the difference in the body shape (e.g., deformation in the head and torso), which indicates the potential discrimination information for gait recognition. Even if the subtle yet informative inter-subject variations are washed out in the registered gait features, they may still remain in the deformation patterns themselves and hence we may get a correct match by taking the deformation patterns into account in a matching stage.

We therefore propose a unified CNN framework, named **DeformGait**, for robust gait recognition against posture changes. For this purpose, we make the most use of the deformation field between a pair of gait features, i.e., we use it not only for registration but also as an essential cue to discriminate the pairs of the same subjects from those of different subjects. More specifically, given a pair of GEIs as an input, a deformable registration network estimates the deformation field between the pair of GEIs, and then generates a pair of registered GEIs. The registered GEIs as well as the deformation field itself are both fed into recognition networks to output discriminative features, respectively, and then they are jointly used to get a recognition result finally. The contributions of this work are three-fold.

1. The deformation pattern between a pair of gait features is used as a recognition cue for the first time.

While the existing deformable registration-based gait recognition works use the deformation patterns to register a pair of gait features, the proposed method also uses the deformation pattern as a cue to distinguish pairs of the same subject from those of different subjects for the first time in the gait recognition community to the best of our knowledge. Thanks to this, we can learn discriminative features both from a pair of registered gait features and the deformation pattern itself.

2. A unified CNN framework combining deformable registration, feature discrimination learning, and deformation discrimination learning.

The proposed DeformGait is a unified CNN framework that combines a state-of-the-art deformable registration module, i.e., pairwise spatial transformer (PST) [46], with a feature-based recognition network (FRN) and a deformation-based recognition network (DRN), which extracts discriminative features from a pair of registered gait features and the deformation pattern, respectively. The whole network is trained in an end-to-end way so as to ensure the optimal recognition accuracy, which achieves a good trade-off between the feature registration effects and the discrimination capability of the deformation patterns.

3. State-of-the-art performance for cross-speed gait recognition.

As speed variation is a typical covariate causing the posture changes, we evaluated the proposed method on the OU-ISIR Gait Database, Treadmill Dataset A [24], which contains both walking and running modes with the largest speed variations. Compared with the existing deformable registration-based method and other benchmarks, the proposed method achieves the state-of-the-art performance in both identification and verification scenarios.

2. Related Work

2.1. Metric learning-based gait recognition

Various metric learning algorithms have been applied to gait recognition to extract covariate-invariant features for better recognition performance. Han and Bhanu [9] applied principal component analysis (PCA) to reduce the dimension of GEIs and subsequent linear discriminant analysis (LDA) to extract discriminative features. Many other techniques have also been incorporated, such as rank support vector machine [28], general tensor discriminant analysis [37], canonical correlation analysis [1, 19], mutual subspace method (MSM) [13, 12], and random subspace method (RSM) [7, 6].

2.2. CNN-based gait recognition

Similar to other computer vision fields, CNN-based methods have also achieved significant performance improvement

in gait recognition. GEIs are often adopted as inputs of the networks with different architectures [49, 31, 43, 34, 48, 21, 50]. For examples, while GEINet [31] uses a single GEI as its input for identity classification, the method [43] uses a pair of GEIs as its input to learn discriminative features and have been demonstrated to be more effective than the single-input ones. Different network structures were suggested depending on recognition tasks (i.e., identification and verification) in [34], which consider the essential difference between these two tasks. Rather than GEIs, raw silhouette images [42] and spatio-temporal features [41] have also been investigated as the network inputs. Chao et al. [4] proposed GaitSet by regarding gait as a set of frames, which obtains the state-of-the-art performance. Zhang et al. [51] used the latent pose features for recognition, which is disentangled from the inputted RGB images.

The CNN-based methods generally yield superior performance than the aforementioned traditional metric learning-based methods. They, however, may wash out subtle inter-subject body shape variations by max-pooling layers.

2.3. Deformable registration-based gait recognition

Several works on gait recognition utilized rigid transformation to cope with speed variations, such as linear stride normalization (SN) for double-support phases [35], Procrustes shape analysis [16, 17], and a factorization-based speed transformation model (STM) for leg joint angles [27]. Moreover, FFD-based geometric transformation models, which can handle non-rigid deformation in a more flexible way, were employed for cross-view gait recognition [5] and cross-speed gait recognition [45]. While the FFD model was used to register a pair of GEIs between different views in [5], it was used to register a pair of single-support GEIs (SSGEIs) [44] between walking and running in [45], which are somewhat invariant against speed variation within the same gait mode (i.e., within-walking and within-running).

While most of the above-mentioned approaches designed subject-independent deformation models, i.e., models common for all the subjects [5, 45], subject-dependent deformation models were employed in a few works. A subject-dependent eigen FFD was constructed by applying an eigenspace method, PCA, to a set of intra-subject deformation fields in [23], which is served as a pre-registration step combined with a recognition network [43]. A subject-dependent pairwise spatial transformer network (PSTN) was presented in [46], which combines a PST for cross-view feature registration and a recognition network for discrimination learning of registered features in a unified training manner.

The aforementioned deformable registration-based methods only consider using the registered features for discrimination learning and matching, which ignores the discrimination capability that is potentially included in the registration deformation fields, i.e., the deformation patterns.

3. DeformGait

3.1. Overview

Similarly to many CNN-based gait recognition works, we adopt a GEI as the network input, which is a frequently used gait feature obtained by averaging normalized silhouette images over a detected gait period [25]. Both the static (e.g., head and torso) and dynamic (e.g., motion in legs and arms) patterns are aggregated in the GEI, which are represented with different intensities, thus resulting in its simple yet effective properties.

As shown in Fig. 2, the proposed DeformGait contains three components: deformable registration module (i.e., PST), FRN, and DRN. Given a matching pair of probe and gallery GEIs, the PST first estimates a non-rigid deformation to register both the probe and gallery GEIs into an intermediate posture, which avoids unnecessary large distortion compared with direct deformation from the source to target posture [46]. The deformation is represented by a vector $\boldsymbol{u}$, which is a set of displacement vectors on the lattice-type control points that indicates the deformation pattern between this pair of inputs (see Fig. 1). The pair of the registered GEIs is then fed into the FRN to learn a discriminative feature representation. The deformation vector $\boldsymbol{u}$ is converted into two 2D matrices of displacements in the horizontal and vertical directions, respectively (visualized as u_h and u_v in Fig. 2), which are further used as the input of DRN for discrimination learning. The output features of FRN and DRN are then concatenated to obtain the final dissimilarity score, which is also used for computing the recognition loss in the training phase.

3.2. Deformable registration via PST

The PST [46] is a spatial transformation module that can be inserted into a recognition network for feature registration, and is derived from spatial transformer network (STN) [14]. During the training of the entire network, the PST is jointly optimized with the loss of the following recognition network, which aims to learn an appropriate deformation for the optimal recognition performance. We will briefly describe it and may refer the readers to [46] for more details.

Given the input pair of probe and gallery GEIs $P^o, G^o \in \mathbb{R}^{H\times W}$, where H and W are the image height and width, respectively, n_h and n_v control points are first evenly allocated along the horizontal and vertical directions on the original images, respectively. The PST then takes the difference image of the original pair as the input, and regresses a FFD-based subject-dependent deformation with the network structure shown in Table 1, which can be formulated as

$$\boldsymbol{u} = T(P^o, G^o), \tag{1}$$

where T denotes the PST, and $\boldsymbol{u} \in \mathbb{R}^{n_h\times n_v\times 2}$ is the deformation parameter vector representing a set of 2D spatial

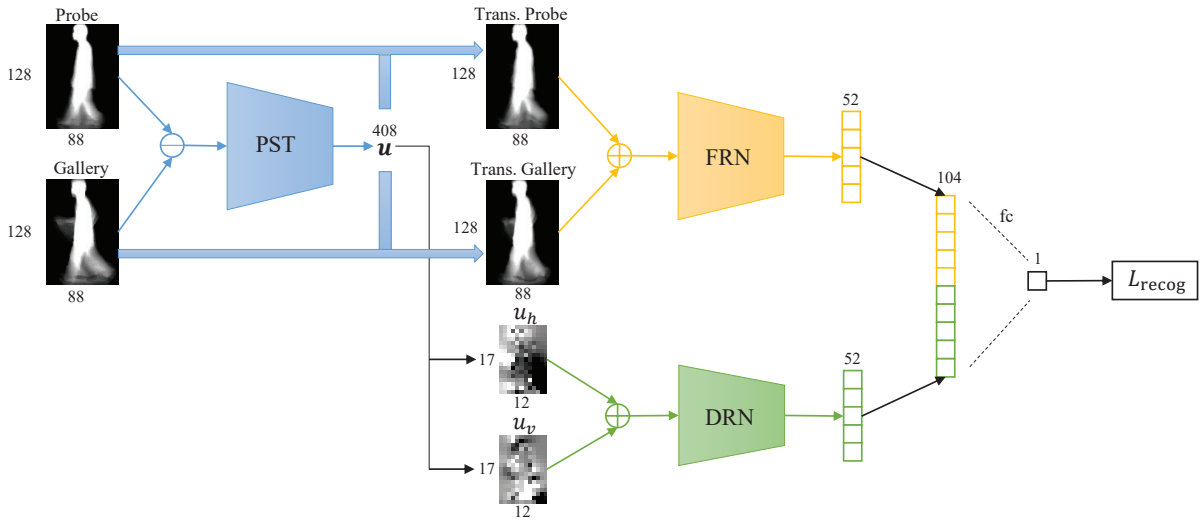

Figure 2. Overview of the proposed DeformGait, which contains a PST, FRN and DRN. The digits indicate the dimensions of the input and output features of each component, and fc denotes the fully connected layer.

displacement vectors on all the control points.

Unlike the traditional STN, the deformation estimated by the PST registers both of the input GEIs into their intermediate state (i.e., posture in our case), which ensures the symmetric deformation between the intermediate posture to each probe and gallery posture. That is, we define $\boldsymbol{u}$ as the deformation between the probe and the intermediate posture, and then, $-\boldsymbol{u}$ is that between the gallery and the intermediate posture. To obtain the registered images, a warping field with the displacements for all pixels, $F(\boldsymbol{u})$, is generated from the deformation vector $\boldsymbol{u}$ using piecewise linear interpolation, and the opposite version $F(-\boldsymbol{u})$ is also obtained similarly. Thus, the registered GEI pair P^t and G^t are sampled using the warping fields as

$$P^t = P^o \circ F(\boldsymbol{u}), \qquad G^t = G^o \circ F(-\boldsymbol{u}), \tag{2}$$

where $\circ$ indicates a deformation operator.

3.3. Discrimination learning with FRN and DRN

3.3.1 FRN

The pair of registered probe and gallery GEIs is then fed into the FRN to be projected into a discriminative space. We employ a network that has a similar structure to a state-of-the-art GEI-based recognition network, LB [43], for the FRN. More specifically, given the deformed GEI pair, the FRN first computes the pixel-wise weighted sum using the paired convolutional filters at the bottom layer, which allows the network to be more sensitive to the local differences, where the inter-subject variations is larger than the intra-subject variations, and hence it is suitable for the input pair that has been registered [23]. Afterwards, an M-dimensional image-based discriminative feature $\boldsymbol{f}^i \in \mathbb{R}^M$ is obtained using the network layers shown in Table 1, which is represented as

$$\boldsymbol{f}^i = f_{\text{FR}}(P^t, G^t), \tag{3}$$

where f_{FR} denotes the FRN, and M is set to 52 in our implementation.

3.3.2 DRN

To learn the discriminative information from the deformation pattern estimated by the PST, we first separate the deformation vector $\boldsymbol{u}$ into two 2D matrices $u_h, u_v \in \mathbb{R}^{n_h \times n_v}$, which are composed of the horizontal and vertical displacements on the control points, respectively. The 2D matrices are then fed into the DRN similarly to the FRN. Because the size of u_h and u_v is much smaller compared with P^t and G^t (i.e., 17×12 vs. 128×88), we use fewer convolutional layers with smaller filter size, and remove the max pooling layer for the DRN, as shown in Table 1. Consequently, an M-dimensional deformation-based discriminative feature $\boldsymbol{f}^d \in \mathbb{R}^M$ is obtained as

$$\boldsymbol{f}^d = f_{\text{DR}}(u_h, u_v), \tag{4}$$

where f_{DR} denotes the DRN.

3.3.3 Combining FRN and DRN

Finally, the dissimilarity between this matching pair is computed using the fusion of the image-based and deformation-based discriminative features. We therefore concatenate the features $\boldsymbol{f}^i$ learned from the FRN, and $\boldsymbol{f}^d$ learned from the DRN, and then obtain the final dissimilarity score s from a $2M$-dimensional concatenated vector using a fully connected layer (denoted by f_{fc}), which is formulated as

$$s = \|f_{\text{fc}}(f_{\text{cat}}(\boldsymbol{f}^i, \boldsymbol{f}^d))\|_1, \tag{5}$$

Table 1. Network architecture of each component. ReLU, LRN, and D indicate the rectified linear unit (ReLU) activation function [29], local response normalization (LRN) [15], and dropout technique [32], respectively.

Module	Layers	Filters/Stride	Acti.	Norm	Max pool
PST	Conv1	16x7x7/1	ReLU	LRN	2x2/2
	Conv2	64x7x7/1	ReLU	LRN	2x2/2
	Fc3	-	ReLU	-	-
	Fc4	-	-	-	-
FRN	Conv1	16x7x7/1	ReLU	LRN	2x2/2
	Conv2	64x7x7/1	ReLU	LRN	2x2/2
	Conv3 (D)	256x7x7/1	ReLU	-	-
	Fc4	-	ReLU	-	-
DRN	Conv1	16x3x3/1	ReLU	LRN	-
	Conv2 (D)	64x3x3/1	ReLU	-	-
	Fc3	-	ReLU	-	-

where f_{cat} denotes a concatenation function.

The output dissimilarity score is also used to calculate the training losses, which are introduced in Section 3.4.

3.4. Recognition task-dependent loss functions

Generally, we consider two tasks for gait recognition, that is, verification (one-to-one-matching), and identification (one-to-many matching). As suggested in [34], we use different loss functions based on the recognition tasks.

In the verification task, a pair of probe and gallery features is given to evaluate whether they have the same identity, which is decided by the comparison between their dissimilarity score and an acceptance threshold. To get a correct verification result, the absolute dissimilarity of the same subject pair is required to be lower than those of the different subject pair. Therefore, the contrastive loss is adopted as the recognition loss for the proposed network in the verification scenario, which is defined as [8]

$$L_{\mathrm{recog}}^{\mathrm{cont}} = \frac{1}{2N}\sum_{n=1}^{N}(\delta_n s_n^2 + (1-\delta_n)\max(\mathrm{margin}-s_n, 0)^2), \quad (6)$$

where N denotes the number of training input pairs, s_n is the obtained dissimilarity score of the n-th input pair, and δ_n is the Kronecker delta for the n-th pair, which is set to one for the same subject pair, and zero for different subject pairs.

In the identification task, a probe is matched with all the galleries to find the one that has the same identity as the probe, which is executed by selecting the gallery with the lowest dissimilarity to the probe. To achieve a correct match for the identification task, the dissimilarity score between the probe and the same subject in the galleries (i.e., genuine) needs to be relatively lower than those between the probe and different subjects (i.e., imposters). Therefore, following [34], we adopt a triplet of probe, genuine and imposter as the input in the identification scenario. More specifically, a triplet input constitutes a genuine pair and an imposter pair, each of which is fed into the proposed network, respectively

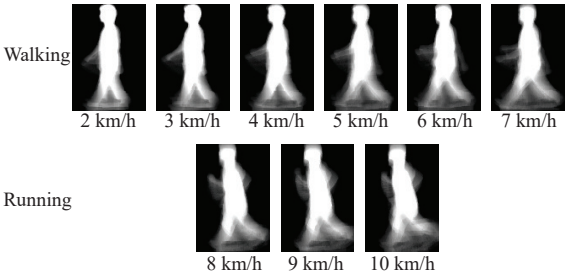

Figure 3. GEI examples from the same subject at different speeds in OUTD-A.

(i.e., parallel networks with shared parameters). The output dissimilarity scores of both the genuine and imposter pairs are used to compute the triplet loss as [40]

$$L_{\mathrm{recog}}^{\mathrm{trip}} = \frac{1}{2N}\sum_{n=1}^{N}\max(\mathrm{margin} - s_n^{\mathrm{imp}} + s_n^{\mathrm{gen}}, 0)^2, \quad (7)$$

where N denotes the number of training input triplets, s_n^{gen} is the dissimilarity score of the genuine pair in the n-th input triplet, and s_n^{imp} is that of the corresponding imposter pair.

4. Experiments

4.1. Dataset

Because the posture changes significantly when the walking speed changes, we evaluated the proposed method on the OU-ISIR Gait Database, Treadmill Dataset A (OUTD-A) [24], which is the dataset with the largest speed variations captured from the side view. It contains 34 subjects with 9 different speeds, ranging from 2 km/h to 10 km/h in 1 km/h interval. This dataset is also the only dataset that includes the image sequences in a running gait mode. More specifically, the speeds from 2 km/h to 7 km/h are considered as the walking mode, while that from 8 km/h to 10 km/h are regarded as the running mode, as examples shown in Fig. 3. We followed the protocol of this dataset [24], that is, 9 subjects were used for training, and the other disjoint 25 subjects were used for testing. In the training phase, the GEIs of all 9 speeds were fed into the proposed network simultaneously to train a single CNN model; in the test phase, following the criteria in [45], the recognition performance is separately evaluated for three cases, within-walking, within-running, and cross-mode (i.e., the probes and galleries have different walking/running modes), whereas the last case appears much greater posture variations compared with the former two cases.

4.2. Implementation details

We trained the proposed network using the stochastic gradient descent algorithm [2] with a batch size of 600. The margin in Eqs. 6 and 7 were both set to 3, and the probability for the dropout technique is set to 0.5.

Table 2. Comparison of the proposed method and other benchmarks for three modes. Bold and bold italic indicate the best and second-best results, respectively. This font convention is used to indicate performance throughout this paper.

(a) Overall rank-1 identification rates (%) for each mode.

Method	Within-walking	Within-running	Cross-mode
DCM [17]	92.4	-	-
RSM [7]	98.1	99.2	53.4
MSM [13]	96.8	-	-
MSM-DA [12]	**99.8**	-	-
SSGEI [45]	***99.3***	**100**	***81.0***
LB [43]	93.3	98.2	70.2
PSTN [46]	95.4	**100**	77.8
DeformGait (proposed)	97.6	**100**	**82.6**

(b) Overall EER (%) for each mode.

Method	Within-walking	Within-running	Cross-mode
SSGEI [45]	***3.1***	3.0	12.9
LB [43]	4.9	2.7	11.6
PSTN [46]	3.5	***2.6***	***9.5***
DeformGait (proposed)	**2.8**	**1.5**	**8.7**

We applied a similar training strategy to PSTN [46]. More specifically, to first achieve relatively stable deformation effects, the PST was first pre-trained using only the same subject pairs of the whole training set with the initial learning rate of 0.001, which was optimized by the L2 loss and a smoothness term [23] to minimize the differences between the transformed GEI pairs output from the PST. Then, the network parameters of the PST in the whole DeformGait were initialized using the pre-trained PST, and the whole DeformGait including all three components was optimized via a unified training manner, using only the recognition losses introduced in Section 3.4 with both the same and different subject pairs. The initial learning rates were set to 0.001 for the PST, and 0.01 for the FRN and DRN, and were reduced by 0.1 four times during the whole training process. Because the number of training subjects (i.e., 9) is insufficient for training a deep CNN from scratch, we fine-tuned the network from the model pre-trained on the largest cross-view gait dataset, i.e., the OU-ISIR Gait Database, Multi-View Large Population Dataset (OU-MVLP) [33], which contains 10,307 subjects with 14 wide view variations.

The recognition performance was evaluated using the receiver operating characteristic (ROC) curve and equal error rate (EER) of the false acceptance rate (FAR) and false rejection rate (FRR) for the verification task, and using the cumulative matching characteristics (CMC) curve and rank-1 identification rate for the identification task.

4.3. Comparison with state-of-the-art methods

We compared the proposed method with the state-of-the-art methods of cross-speed gait recognition, i.e., the differential composition model (DCM) [17], RSM [7], MSM [13], MSM using divided area (MSM-DA) [12], SSGEI [45], and the state-of-the-art CNN-based method, i.e., LB [43], in addition to the state-of-the-art deformable registration-based method, i.e., PSTN [46][1]. Results are shown in Table 2. Because the first four methods in Table 2(a) did not provide the EERs, we only compared with the other three ones for verification task in Table 2(b). We also reported the rank-1 identification rates of all 81 speed combinations for the proposed method in Table 3.

Because of the quite limited training data, the overfitting problem often occurs for the CNN-based methods, and hence the LB and PSTN do not perform better than the traditional methods. Compared with the LB and PSTN, the proposed method achieves better results for all three evaluation cases, which shows the effectiveness of considering both feature registration and deformation discrimination learning for recognition. Although the proposed method obtains slightly worse results for the within-walking case in Table 2(a), it is worth mentioning to that MSM-DA only focuses on the static gait patterns, and hence is unsuitable for the cross-mode case. Additionally, the extraction of SSGEI, which is a specially designed gait feature against speed variations, and the machine learning techniques used in [45], were executed for each of the three cases, respectively, while the proposed method trained a single unified CNN model applicable for all three cases simultaneously.

On the other hand, the proposed DeformGait yields the best result for the cross-mode case, which is much more challenging because the postures in both static and dynamic body parts vary between the walking and running modes, whereas the within-walking and running cases mainly involve the posture changes in dynamic parts. As shown in Table 2(b), the proposed method performs better than others for all cases in the verification scenario. Therefore, in general, the proposed method achieves state-of-the-art performance.

4.4. Ablation study

We analyzed the effects of each component of the proposed method in Table 4. The first row shows the results of baseline FRN, where the features with speed (i.e., posture) variations were directly used for feature learning without deformable registration. In the second row, the DRN was removed from the proposed DeformGait, i.e., the recognition was only based on the registered features; in the third row, the FRN was removed, i.e., the recognition was only based on the deformation patterns. The last row is the results of the whole proposed framework. In Fig. 4, we also showed the ROC and CMC curves for a challenging case in the within-walking (i.e., 2 km/h vs. 7 km/h) and cross-mode (i.e., 2 km/h vs. 10 km/h) scenario, respectively.

[1]Results of LB are from [45], and the PSTN was also fine-tuned from the pre-trained model on OU-MVLP for a fair comparison.

Table 3. Rank-1 identification rates (before slash) and EER (after slash) [%] of the proposed method for all 81 speed combinations. Probe and gallery are denoted as P and G, respectively.

P \ G	2km/h	3km/h	4km/h	5km/h	6km/h	7km/h	8km/h	9km/h	10km/h
2km/h	100/1.2	100/0.3	96/0.5	84/4.0	88/2.7	96/3.2	68/7.5	76/10.0	72/8.0
3km/h	100/4.0	100/0.0	96/0.7	92/4.0	96/2.7	96/2.8	76/8.0	80/8.0	68/8.0
4km/h	100/2.5	100/0.8	100/1.0	100/4.0	96/3.5	92/3.0	76/9.2	84/9.2	76/10.5
5km/h	100/1.7	100/0.3	100/0.2	100/4.0	96/4.0	96/4.0	76/11.5	88/10.0	80/10.3
6km/h	100/3.3	100/4.0	100/4.0	100/8.0	100/1.3	100/1.7	80/5.7	96/5.5	72/5.7
7km/h	92/3.5	96/4.3	100/4.0	100/7.0	100/0.3	100/0.2	84/5.2	92/8.0	84/5.3
8km/h	76/7.8	80/6.0	96/8.2	92/8.2	100/5.2	92/4.8	100/1.5	100/1.5	100/1.8
9km/h	80/8.0	84/7.3	96/10.7	84/10.8	100/8.0	96/5.7	100/0.8	100/1.0	100/0.2
10km/h	60/10.2	84/9.8	84/15.2	80/15.0	80/8.7	80/8.0	100/0.8	100/4.0	100/0.2

Table 4. Overall rank-1 identification rates (before slash) and EER (after slash) [%] of each mode for ablation experiments.

Method	Within-walking	Within-running	Cross-mode
FRN	94.8/3.6	99.1/2.6	71.8/13.5
PST-FRN	95.4/3.5	**100**/2.6	77.8/9.5
PST-DRN	91.6/6.9	98.7/6.2	60.3/16.8
PST-FRN-DRN (proposed)	**97.6/2.8**	**100/1.5**	**82.6/8.7**

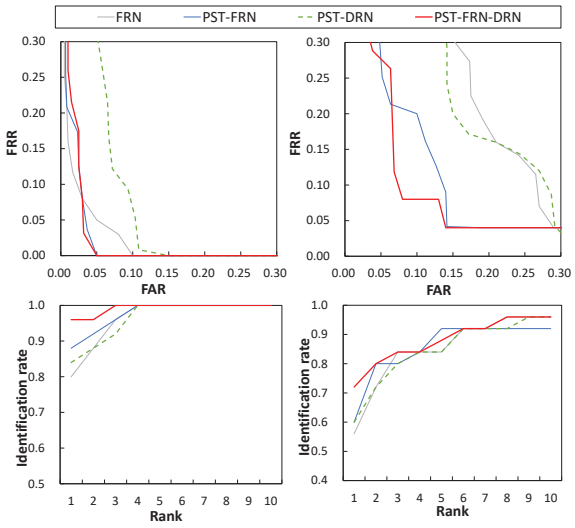

Figure 4. ROC (top) and CMC (bottom) curves for ablation experiments (left: 2 km/h probe vs. 7 km/h gallery, right: 2 km/h probe vs. 10 km/h gallery).

Comparing the results of FRN and PST-FRN, the deformable registration clearly improves the performance, especially for the difficult cross-mode case. While the registered gait features (i.e., GEI) contain both shape and motion information, where the intra-subject variations have been mitigated by the PST, the deformation patterns just include relatively low-dimensional spatial displacements, and hence it is insufficient to use only the deformation patterns for discriminating subjects, which is demonstrated by the performance gap between PST-FRN and PST-DRN. However, by combining the DRN with FRN, where the learned discriminative features from the deformation patterns are counted as additional information to help with the possible failures by the FRN, the performance of the proposed method were further improved, which achieved the best results for both verification and identification tasks. Therefore, all three components contributed to the proposed method as a result.

For most of the methods, the results of within-running case have almost saturated, since the posture variations are relatively smaller among different running speeds, which are more reflected in the changes of gait periods (e.g., shorter gait period for larger running speed) [45]. The within-walking case is relatively more challenging than the within-running, as the arm swing and stride vary more obviously with the change of walking speeds. The cross-mode case is the most difficult among the three cases, where the posture changes exist in both the static (e.g., bending the upper body forward when changing from walking to running) and dynamic (e.g., flexing knees and bending arms for running poses) body parts (see Fig. 3); hence, the performance differences are more obvious for the cross-mode case, which shows the effectiveness of deformable registration and deformation-based discrimination learning for tackling posture variations.

4.5. Discussion

We first show a successful matching example in Fig. 5(1) to illustrate the effectiveness of the proposed method. We choose the most difficult cross-mode case, i.e., the probe is in 2 km/h while two galleries are in 10 km/h, which contain a genuine subject and an imposter subject, respectively.

Due to the obvious posture variations caused by speed change, the dissimilarity between the original genuine pair is considerably large (see Fig. 5(1-c)). On the other hand, the running imposter subject does not bend the upper body forward like most subjects, and hence the posture differences between the imposter pair are much smaller than that between the genuine pair.

Because registration may reduce both the intra-subject

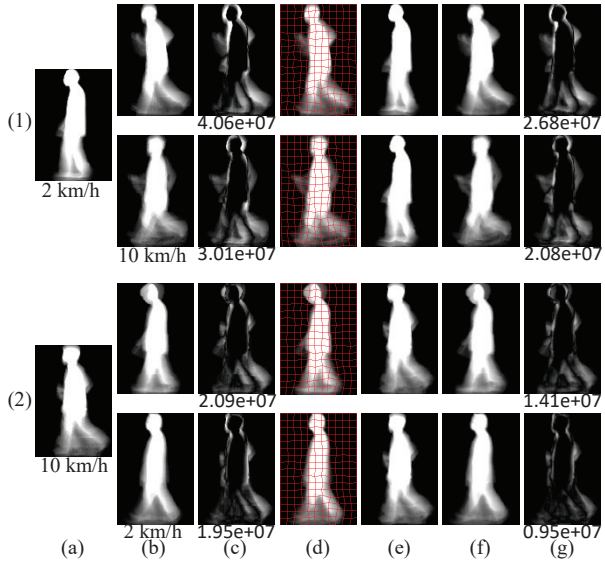

Figure 5. Examples of the deformations and registered features by the proposed method. (1) A success case; (2) a failure case. In each case, first row shows the results of the genuine, and second row shows that of the imposter. (a) Original probe GEI. (b) Original gallery GEI. (c) Difference image and dissimilarity score between (a) and (b). (d) Deformation field for the gallery learned by the DeformGait. (e) Registered probe. (f) Registered gallery. (g) Difference image and dissimilarity score between (e) and (f).

variations and inter-subject variations, the PST module, which is supervised by the recognition loss, works in a way that reduces the intra-subject variations caused by the posture changes, while not overly registering the imposter pairs to maintain the inter-subject variations [46]. Therefore, even the posture differences between the genuine pair are just somewhat reduced by deforming both of them into a intermediate posture (e.g., the upper body of the genuine becomes relatively straight (see Figs. 5(1-b) and (1-f))). However, the dissimilarity between the imposter pair is still smaller than that of the genuine pair because of their quite similar posture in the upper body; hence using only the registered features resulted in an incorrect match (see Fig. 5(1-g)).

On the other hand, we can observe that there are clear differences between the deformation fields of the genuine pair and imposter pair. For example, the deformation pattern around the neck and back of the genuine has an effect of stretching backwards to make the upper body somewhat straight; while a shrink effect exists in the deformation pattern around the stomach and hip of the imposter, which is resulted from the body shape difference between the probe and imposter subjects (see Fig. 5(1-d)). By taking the differences among the deformation patterns into account, the proposed method finally yielded a correct match by a dissimilarity score of 1.78 and 8.35 for the genuine pair and imposter pair, respectively, which demonstrates the effectiveness of considering discriminations in the deformation patterns for recognition.

We next analyze a typical failure case in Fig. 5(2). Because the genuine temporarily changes his posture (i.e., look down) within a gait period, which blurs the head part in the GEI (see Fig. 5(2-b)), and hence causes relatively large intra-subject difference between the genuine pair. This large dissimilarity is not effectively reduced even after registration, since the problem caused by blurred body parts cannot be well solved via the deformable registration (see Fig. 5(2-f)).

To register the temporary posture variations, the deformation field for the genuine tries pulling the head part backward (Fig. 5(2-d)), which does not usually appear in the registration for a walking genuine to a running probe (the head part is usually deformed to slightly bend forward to resemble a running pose). On the other hand, the imposter does not have such temporary posture change and also owns a similar body shape to the probe subject, and hence the deformation displacements for the imposter is small, which is more likely to be that for the same subject pair. Therefore, the proposed method yielded a false match despite that we consider both the registered features and deformation patterns.

One possible solution to this problem is to consider the deformable registration for each frame, rather than applying that for the GEI that may contain the blurred region. Additionally, since the temporary posture change occurs for only several frames in a period, the effects of large deformation displacements for these frames can be mitigated if we consider discrimination learning using deformation patterns for all frames, which is beneficial for matching task.

5. Conclusion

This paper presented a method of robust gait recognition against posture changes, named DeformGait. Given an input matching pair of GEIs, the PST first registered both of them to mitigate the posture variations via the learned subject-dependent deformation. The registered GEI pair and the deformation pattern were then fed into the FRN and DRN for discrimination learning, respectively, which were further combined to obtain the final dissimilarity score. Experiments illustrated the state-of-the-art performance of the proposed method for cross-speed gait recognition.

One future research goal is to extend the proposed method considering registration for each frame to tackle the temporary posture changes within a period. Another future work is to evaluate the performance of the proposed method for other covariates, such as cross-view gait recognition.

Acknowledgment. This work was supported by JSPS KAKENHI Grant No. JP18H04115, JP19H05692, and JP20H00607, Jiangsu Provincial Science and Technology Support Program (No. BE2014714), the 111 Project (No. B13022), and the Priority Academic Program Development of Jiangsu Higher Education Institutions.

References

[1] K. Bashir, T. Xiang, and S. Gong. Cross view gait recognition using correlation strength. In *BMVC*, 2010.

[2] L. Bottou and O. Bousquet. The tradeoffs of large scale learning. In *Proceedings of the 20th International Conference on Neural Information Processing Systems*, NIPS'07, pages 161–168, USA, 2007. Curran Associates Inc.

[3] I. Bouchrika, M. Goffredo, J. Carter, and M. Nixon. On using gait in forensic biometrics. *Journal of Forensic Sciences*, 56(4):882–889, 2011.

[4] H. Chao, Y. He, J. Zhang, and J. Feng. Gaitset: Regarding gait as a set for cross-view gait recognition. In *Proc. of the 33th AAAI Conference on Artificial Intelligence (AAAI 2019)*, 2019.

[5] H. El-Alfy, C. Xu, Y. Makihara, D. Muramatsu, and Y. Yagi. A geometric view transformation model using free-form deformation for cross-view gait recognition. In *2017 4th IAPR Asian Conference on Pattern Recognition (ACPR)*, pages 929–934, 2017.

[6] Y. Guan, C. Li, and F. Roli. On reducing the effect of covariate factors in gait recognition: A classifier ensemble method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(7):1521–1528, July 2015.

[7] Y. Guan and C.-T. Li. A robust speed-invariant gait recognition system for walker and runner identification. In *Proc. of the 6th IAPR International Conference on Biometrics*, pages 1–8, 2013.

[8] R. Hadsell, S. Chopra, and Y. LeCun. Dimensionality reduction by learning an invariant mapping. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, pages 1735–1742, June 2006.

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

[10] Y. He, J. Zhang, H. Shan, and L. Wang. Multi-task gans for view-specific feature learning in gait recognition. *IEEE Transactions on Information Forensics and Security*, 14(1):102–113, Jan 2019.

[11] H. Iwama, D. Muramatsu, Y. Makihara, and Y. Yagi. Gait verification system for criminal investigation. *IPSJ Transactions on Computer Vision and Applications*, 5:163–175, Oct. 2013.

[12] Y. Iwashita, M. Kakeshita, H. Sakano, and R. Kurazume. Making gait recognition robust to speed changes using mutual subspace method. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2273–2278, 2017.

[13] Y. Iwashita, H. Sakano, and R. Kurazume. Gait recognition robust to speed transition using mutual subspace method. In *Image Analysis and Processing — ICIAP 2015: 18th International Conference, Genoa, Italy, September 7-11, 2015, Proceedings, Part I*, pages 141–149, Cham, 2015.

[14] M. Jaderberg, K. Simonyan, A. Zisserman, and k. kavukcuoglu. Spatial transformer networks. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28*, pages 2017–2025. Curran Associates, Inc., 2015.

[15] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei. Large-scale video classification with convolutional neural networks. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1725–1732, June 2014.

[16] W. Kusakunniran, Q. Wu, J. Zhang, and H. Li. Speed-invariant gait recognition based on procrustes shape analysis using higher-order shape configuration. In *The 18th IEEE Int. Conf. Image Processing*, pages 545–548, 2011.

[17] W. Kusakunniran, Q. Wu, J. Zhang, and H. Li. Gait recognition across various walking speeds using higher order shape configuration based on a differential composition model. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 42(6):1654–1668, Dec. 2012.

[18] W. Kusakunniran, Q. Wu, J. Zhang, and H. Li. Gait recognition under various viewing angles based on correlated motion regression. *IEEE Transactions on Circuits and Systems for Video Technology*, 22(6):966–980, 2012.

[19] W. Kusakunniran, Q. Wu, J. Zhang, H. Li, and L. Wang. Recognizing gaits across views through correlated motion co-clustering. *IEEE Transactions on Image Processing*, 23(2):696–709, Feb 2014.

[20] S. Lee, Y. Liu, and R. Collins. Shape variation-based frieze pattern for robust gait recognition. In *Proc. of the 2007 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, pages 1–8, Minneapolis, USA, Jun. 2007.

[21] X. Li, Y. Makihara, C. Xu, Y. Yagi, and M. Ren. Joint intensity transformer network for gait recognition robust against clothing and carrying status. *IEEE Transactions on Information Forensics and Security*, 14(12):3102–3115, Dec 2019.

[22] N. Lynnerup and P. Larsen. Gait as evidence. *IET Biometrics*, 3(2):47–54, 6 2014.

[23] Y. Makihara, D. Adachi, C. Xu, and Y. Yagi. Gait recognition by deformable registration. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018.

[24] Y. Makihara, H. Mannami, A. Tsuji, M. Hossain, K. Sugiura, A. Mori, and Y. Yagi. The ou-isir gait database comprising the treadmill dataset. *IPSJ Transactions on Computer Vision and Applications*, 4:53–62, Apr. 2012.

[25] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, and Y. Yagi. Gait recognition using a view transformation model in the frequency domain. In *Proc. of the 9th European Conference on Computer Vision*, pages 151–163, Graz, Austria, May 2006.

[26] Y. Makihara, A. Suzuki, D. Muramatsu, X. Li, and Y. Yagi. Joint intensity and spatial metric learning for robust gait recognition. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6786–6796, July 2017.

[27] Y. Makihara, A. Tsuji, and Y. Yagi. Silhouette transformation based on walking speed for gait identification. In *Proc. of the 23rd IEEE Conf. on Computer Vision and Pattern Recognition*, San Francisco, CA, USA, Jun 2010.

[28] R. Martin-Felez and T. Xiang. Uncooperative gait recognition by learning to rank. *Pattern Recognition*, 47(12):3793 – 3806, 2014.

[29] V. Nair and G. E. Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10, pages 807–814, USA, 2010. Omnipress.

[30] T. W. Sederberg and S. R. Parry. Free-form deformation of solid geometric models. *SIGGRAPH Comput. Graph.*, 20(4):151–160, Aug. 1986.

[31] K. Shiraga, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi. Geinet: View-invariant gait recognition using a convolutional neural network. In *2016 International Conference on Biometrics (ICB)*, pages 1–8, 2016.

[32] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 2014.

[33] N. Takemura, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi. Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition. *IPSJ Trans. on Computer Vision and Applications*, 10(4):1–14, 2018.

[34] N. Takemura, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi. On input/output architectures for convolutional neural network-based cross-view gait recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, pages 1–1, 2018.

[35] R. Tanawongsuwan and A. Bobick. Modelling the effects of walking speed on appearance-based gait recognition. *Proc. of the 17th IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2:783–790, 2004.

[36] D. Tao, X. Li, X. Wu, and S. Maybank. Human carrying status in visual surveillance. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, volume 2, pages 1670–1677, New York, USA, Jun. 2006.

[37] D. Tao, X. Li, X. Wu, and S. J. Maybank. General tensor discriminant analysis and gabor features for gait recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(10):1700–1715, Oct 2007.

[38] N. F. Troje. Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of Vision*, 2:371–387, 2002.

[39] C. Wang, J. Zhang, L. Wang, J. Pu, and X. Yuan. Human identification using temporal information preserving gait template. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2164 –2176, nov. 2012.

[40] J. Wang, Y. Song, T. Leung, C. Rosenberg, J. Wang, J. Philbin, B. Chen, and Y. Wu. Learning fine-grained image similarity with deep ranking. In *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '14, pages 1386–1393, Washington, DC, USA, 2014. IEEE Computer Society.

[41] T. Wolf, M. Babaee, and G. Rigoll. Multi-view gait recognition using 3d convolutional neural networks. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 4165–4169, 2016.

[42] Z. Wu, Y. Huang, and L. Wang. Learning representative deep features for image set analysis. *IEEE Transactions on Multimedia*, 17(11):1960–1968, Nov 2015.

[43] Z. Wu, Y. Huang, L. Wang, X. Wang, and T. Tan. A comprehensive study on cross-view gait based human identification with deep cnns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(2):209–226, 2017.

[44] C. Xu, Y. Makihara, X. Li, Y. Yagi, and J. Lu. Speed invariance vs. stability: Cross-speed gait recognition using single-support gait energy image. In *Proc. of the 13th Asian Conf. on Computer Vision (ACCV 2016)*, pages 52–67, Taipei, Taiwan, Nov. 2016.

[45] C. Xu, Y. Makihara, X. Li, Y. Yagi, and J. Lu. Speed-invariant gait recognition using single-support gait energy image. *Multimedia Tools and Applications*, 78:26509–26536, Sep 2019.

[46] C. Xu, Y. Makihara, X. Li, Y. Yagi, and J. Lu. Cross-view gait recognition using pairwise spatial transformer networks. *IEEE Transactions on Circuits and Systems for Video Technology*, pages 1–1, 2020.

[47] D. Xu, S. Yan, D. Tao, L. Zhang, X. Li, and H. jiang Zhang. Human gait recognition with matrix representation. *IEEE Trans. Circuits Syst. Video Technol*, 16(7):896–903, 2006.

[48] S. Yu, H. Chen, E. B. G. Reyes, and N. Poh. Gaitgan: Invariant gait feature extraction using generative adversarial networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 532–539, July 2017.

[49] C. Zhang, W. Liu, H. Ma, and H. Fu. Siamese neural network based gait recognition for human identification. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2832–2836, 2016.

[50] K. Zhang, W. Luo, L. Ma, W. Liu, and H. Li. Learning joint gait representation via quintuplet loss minimization. In *2019 conference on computer vision and pattern recognition (CVPR 2019)*, 2019.

[51] Z. Zhang, L. Tran, X. Yin, Y. Atoum, and X. Liu. Gait recognition via disentangled representation learning. In *2019 conference on computer vision and pattern recognition (CVPR 2019)*, 2019.