

Gait Recognition by Deformable Registration

Yasushi Makihara¹, Daisuke Adachi¹, Chi Xu^{2,1}, Yasushi Yagi¹

¹Osaka University, ²Nanjing University of Science and Technology

{makihara, adachi, xu, yagi}@am.sanken.osaka-u.ac.jp

Abstract

This paper describes a method of gait recognition robust against intra-subject posture changes. A person sometimes walks with changing his/her posture when looking down at a smartphone or carrying a heavy object, which makes intra-subject variation large and consequently makes gait recognition difficult. We therefore introduce a deformable registration model to mitigate the intra-subject posture changes. More specifically, we represent a deformation field by a set of deformation vectors on lattice-type control points allocated on an image, i.e., by free-form deformation (FFD) framework. Given a pair of a probe and a gallery, we compute the deformation field so as to minimize the difference between a probe morphed by the deformation field and the gallery, as well as to ensure the spatial smoothness of the deformation field. We then learn the intra-subject eigen deformation modes from a training set of the same subjects' pairs (e.g., bending the upper body forward and swinging arms more), which are relatively different from inter-subject deformation modes (e.g., body shape spread and stride change). Moreover, because the deformable registration is responsible for a preprocessing part before matching, it can be combined with any types of matching algorithms for gait recognition. Experiments with 1,334 subjects show that the proposed method improves the gait recognition accuracy in both cases without and with a state-of-the-art deep learning-based matcher, respectively.

1. Introduction

Gait recognition [48] is one of behavioral biometrics and has advantages over other physiological biometrics such as face, iris, or finger vein because gait is observable (i) even when a target person is at a distance from a camera since it can be recognized from a relatively low-resolution image sequence [45, 51, 77], and (ii) without subject cooperation [6, 44] since people unconsciously exhibit their own natural walking styles in a daily life. Thanks to these advantages, gait recognition is suitable for many potential ap-

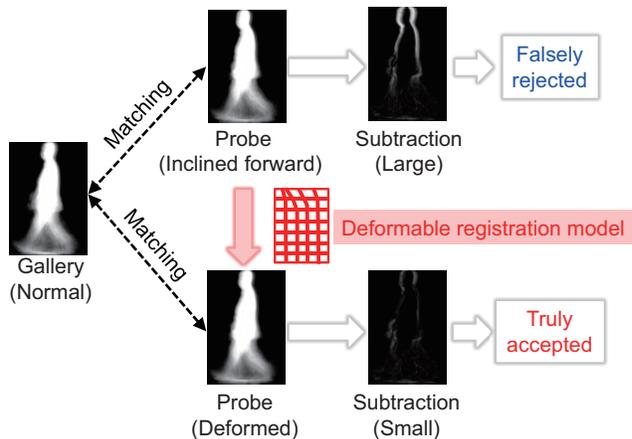


Figure 1: A matching example of the same subject's pair before (top) and after (bottom) the deformable registration. Given a gallery with a normal posture and a probe inclined forward, direct matching may result in false rejection due to a large dissimilarity. The deformable registration can mitigate such a posture change and make the dissimilarity small, which results in true acceptance.

plications such as surveillance, forensics, and criminal investigation [9, 22, 40].

In the gait recognition field, appearance-based gait representations [5, 6, 19, 27, 41, 46, 52, 65, 66] which directly use input or silhouette images in a holistic way to extract gait features without a human model fitting, are widely employed. In particular, silhouette-based representations such as gait energy images (GEIs) [19] are dominant because of their simple yet effective properties. The appearance-based approaches, however, often suffer from large intra-subject appearance changes due to covariates such as clothing [5, 6, 21, 35, 50], carrying status [14, 59, 60], view [26, 28, 30, 32, 39, 41, 54, 55, 69], and walking speed [2, 16, 29, 31, 38, 42, 43, 57].

Despite of the above mentioned extensive studies on robust gait recognition, intra-subject posture changes have not yet fully investigated. People may sometimes walk with doing something else at the same time, e.g., with looking

down to watch a smartphone, with looking up to see natural scenery such as cherry blossom, with bending the upper body forward to carry a heavy backpack. In addition, people may change their postures during walking depending on physical or mental status, e.g., people may stoop when feeling tired and may swing their arms more with the emotion “happy” [62]. As such, human gait contains more or less posture changes, which make the gait recognition difficult as shown in Fig. 1.

A potential strategy to cope with the posture changes is to employ spatial metric learning, which has already been introduced to gait recognition robust against various covariates [17–19, 37, 59, 61, 65, 71, 73]. Roughly speaking, the spatial metric learning aims at assigning low weights to pixels whose intra-subject variation is large (e.g., pixels around his/her back for the posture changes) and vice versa for better recognition accuracy.

The spatial metric learning is, however, not necessarily the optimal way to cope with the posture changes. Assuming that subjects have a variety of body shapes (e.g., a fat or thin torso), spatial positions where the intra-subject posture changes affect, may be substantially different (e.g., more outer and inner sides of the torso for fatter and thinner subjects, respectively). Consequently, low weights are assigned to a wide area around the torso’s contour, and hence body shape information which is useful for discrimination, may be washed out.

A more direct and intuitive way to cope with the posture changes is deformable registration, because a deformation field can effectively represent the posture changes regardless of the body shape variations, unlike the spatial metric learning. For examples, once a deformation field to incline the upper body forward is prepared as shown in Fig. 1, it is commonly applicable to both a fat and thin subjects.

Such deformation models are considered in a few studies on gait recognition. For examples, several studies on cross-speed gait recognition consider stretching the lower body parts (i.e., legs) horizontally at double support phases [58], transformation of joint angle sequences [42], and Procrustes shape analysis [31]. The above mentioned approaches cannot handle non-rigid deformation [31, 58] or suffer from error-prone human model fitting with high computational cost [42]. An exception is a geometric transformation model for cross-view gait recognition [15] which represents non-rigid deformation between two views by a free form deformation framework [53]. The method, however, defines only a single and common deformation field for a specific pair of views in a subject-independent way, and hence cannot represent subject-dependent (i.e., matching pair-dependent) deformation fields.

Looking at other biometrics fields, we notice that subject-dependent facial image deformation (alignment) is often incorporated for cross-view face recognition [24, 25,

70], face recognition robust against facial expressions [8, 20], or facial expression analysis [49], using a standard deformation technique such as active shape model [12]. A key to the successful deformation is landmark detection from outstanding facial parts such as eyes, nose, and mouth. Such outstanding landmarks are, however, unavailable for low-resolution gait images.

We therefore introduce a deformable registration model to cope with the posture changes in gait recognition. The contributions of this work are as follows.

1. A deformation registration model for gait recognition robust against the posture changes.

We introduce a free-form deformation (FFD) framework which enables us to represent more flexible deformation than previous simple deformations [31, 58]. Because we extract eigen deformation modes from a training set of intra-subject deformations, a subject-dependent (i.e., matching pair-dependent) deformation can be expressed by their combination, unlike previous work on cross-view gait recognition [15] just employs a single and fixed deformation mode. Moreover, because the FFD is represented by lattice-type control points, the landmark detection is unnecessary unlike the facial image registration [8, 20, 24, 25, 70].

2. Flexible combination with matching algorithms is possible.

Because the proposed deformable registration model is only responsible for a pre-processing part before matching, it can be combined with any types of metric learning and matching algorithms. We show the effectiveness of the deformable registration model when combined with a state-of-the-art deep learning-based matching algorithm for gait recognition [69].

2. Related work

2.1. Gait representation

Approaches to gait recognition mainly fall into two families: model-based approaches and appearance-based (model-free) approaches. The model-based approaches [3, 7, 13, 34, 63, 64, 72, 79] fit articulated human body models to images and extract kinematic features such as joint angle sequences. While the model-based approaches may have a potential to more directly represent the posture changes (e.g., incline the upper body forward by rotating a torso joint), they have some drawbacks: error-prone fitting and high computational costs.

The majority of appearance-based approaches aggregate a gait image sequence over one gait period (cycle) to generate an image-based gait template such as GEI [19], frequency-domain feature [41], chrono-gait image [65], gait entropy image [5], gait flow image [33], frame difference energy image [11], and Gabor GEIs [61]. The proposed deformable registration model is applicable to any types of

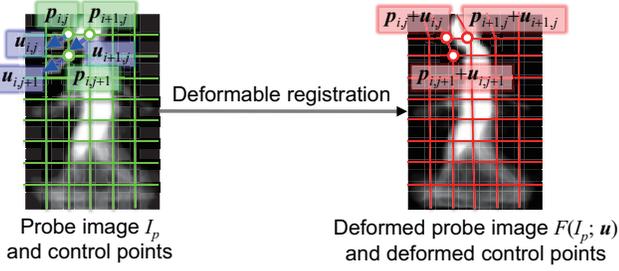


Figure 2: Definition of FFD.

image-based gait templates as above.

2.2. Matching algorithms for gait recognition

When matching gait features, spatial metric learning algorithms are often incorporated, such as linear discriminant analysis [19, 37, 65], general tensor discriminant analysis [59, 61], discriminant analysis with tensor representation [71, 73], or the random subspace method [17, 18].

Moreover, inspired by the great success of deep learning in computer vision, pattern recognition, and also biometrics fields, deep learning-based matching algorithms have been used for gait recognition for several years. Similarly to other recognition tasks such as object recognition, most of them adopt convolutional neural network (CNN) architectures [1, 10, 67–69, 74–76, 78].

As mentioned in the introduction section, because the proposed method is responsible for the pre-processing part, it can be combined with any types of above mentioned matching algorithms for better recognition accuracy.

3. Gait recognition using eigen FFD

3.1. FFD

In this subsection, we introduce a representation of FFD on an image-based gait template such as GEI [19]. Unlike a facial image has outstanding landmarks (e.g., eyes, nose, mouth), a gait template does not have such landmarks. We therefore assign a set of lattice-type control points instead of the landmarks on the image in order to represent the deformation with FFD framework [53] (see Fig. 2).

More specifically, given a probe image I_p with the width W and the height H , we set M_x and M_y lattice-type control points for each of horizontal and vertical direction, with a horizontal interval $\Delta_x = (W - 1)/(M_x - 1)$, and a vertical interval $\Delta_y = (H - 1)/(M_y - 1)$. Note that the number of total control points is $M = M_x M_y$. The spatial position of the control point at the i -th column and j -th row (let it be the (i, j) -th control point for simplicity) is denoted as $\mathbf{p}_{i,j} = [(i - 1)\Delta_x, (j - 1)\Delta_y]^T$ ($i = 1, \dots, M_x, j = 1, \dots, M_y$).

We then introduce a deformation vector $\mathbf{u}_{i,j} \in \mathbb{R}^2$ on the (i, j) -th control point, and define a set of the

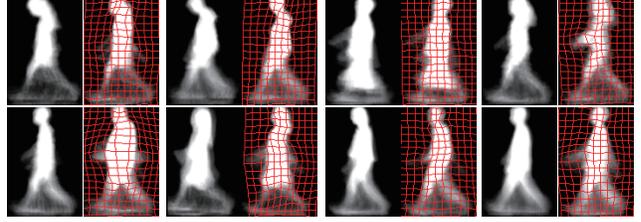


Figure 3: Examples of deformations for the same subject pairs (top) and different subjects pair (bottom). Each pair consists of a probe (left) and a gallery with deformed control points (right).

deformation vectors on all the control points as $\mathbf{u} = [\mathbf{u}_{1,1}^T, \dots, \mathbf{u}_{1,M_y}^T, \dots, \mathbf{u}_{M_x,1}^T, \dots, \mathbf{u}_{M_x,M_y}^T]^T \in \mathbb{R}^{2M}$. Note that the deformation field over all the spatial positions including intermediate points among the control points, is represented by bi-linear interpolation from the deformation vectors at adjacent control points. The (i, j) -th control point is subsequently warped from $\mathbf{p}_{i,j}$ to $\mathbf{p}_{i,j} + \mathbf{u}_{i,j}$, and the probe image I_p is morphed to $\mathcal{F}(I_p; \mathbf{u})$, where \mathcal{F} is a function to map the probe image I_p to a morphed image by the deformation vector \mathbf{u} .

Next, given a pair of the probe image I_p and a gallery image I_g , we estimate the optimal deformation vector from the probe image I_p to the gallery image I_g by minimizing the following objective function:

$$E(\mathbf{u}) = d(\mathcal{F}(I_p; \mathbf{u}), I_g) + \lambda R(\mathbf{u}), \quad (1)$$

where $d(\cdot, \cdot)$ is a function to return sum of absolute difference (SAD) between two images, i.e., l_1 norm, $R(\mathbf{u})$ is a smoothness term to enforce the consistency between deformation vectors on the adjacent control points, and λ is a hyper-parameter to control the strength of the smoothness term. The smoothness term $R(\mathbf{u})$ is defined as

$$R(\mathbf{u}) = \sum_{i=1}^{M_x-1} \sum_{j=1}^{M_y} \|\mathbf{u}_{i+1,j} - \mathbf{u}_{i,j}\|^2 + \sum_{i=1}^{M_x} \sum_{j=1}^{M_y-1} \|\mathbf{u}_{i,j+1} - \mathbf{u}_{i,j}\|^2. \quad (2)$$

3.2. Eigen FFD

If we directly apply the optimal deformation for each matching pair of a probe and a gallery, a dissimilarity measure always becomes small regardless of the same or different subjects' pair, which does not help to improve the gait recognition accuracy. We therefore need to consider the difference of deformation modes between the same and different subjects' pairs.

For this purpose, we observed trends of the deformation fields between the same and different subjects' pairs as shown in Fig. 3. Deformations for the same subjects' pairs mainly come from posture difference (e.g., looking down), slight motion difference (e.g., the degree of arm swing), and

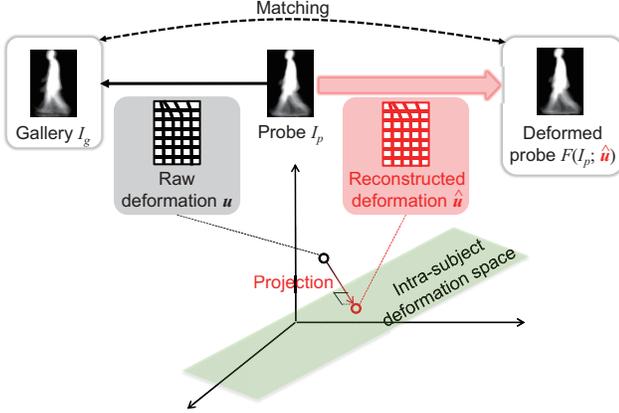


Figure 4: Framework of intra-subject deformation model.

rigid image registration errors when constructing a gait template. On the other hand, those for different subjects' pairs come from body shape difference (e.g., middle-age spread, head-to-body ratio), and more essential motion difference (e.g., almost no arm swing vs. large arm swing, large stride changes).

Because we want to register the same subjects' pairs better to get smaller dissimilarity measures, while keeping the dissimilarity measures of different subjects' pairs large, we constrain deformation modes into a low-dimensional subspace spanned by the intra-subject deformation fields of the same subjects' pairs (see Fig. 4). For this purpose, we train an eigen FFD using intra-subject deformation fields in the training set. More specifically, we apply principal component analysis to the set of intra-subject deformation fields and obtain mean deformation field $\bar{\mathbf{u}} \in \mathbb{R}^{2M}$ and a set of the K largest eigen vectors (deformation modes) $V = [v_1, \dots, v_K] \in \mathbb{R}^{2M \times K}$.

As such, given a raw deformation vector \mathbf{u} between a pair of a probe image I_p and a gallery I_g by minimizing Eq. (1), we obtain a constrained version of the deformation vector $\hat{\mathbf{u}}$ through the eigen FFD as

$$\hat{\mathbf{u}} = VV^T(\mathbf{u} - \bar{\mathbf{u}}) + \bar{\mathbf{u}}. \quad (3)$$

Once we obtain the constrained deformation vector $\hat{\mathbf{u}}$, the probe image I_p is similarly morphed to $\mathcal{F}(I_p; \hat{\mathbf{u}})$ and then the morphed probe image is matched to the gallery image I_g . Specifically, a dissimilarity measure D between them is computed as

$$D = d(\mathcal{F}(I_p; \hat{\mathbf{u}}), I_g). \quad (4)$$

Remind that the function $d(\cdot, \cdot)$ returns SAD of two images as described before.

3.3. Metric learning

Since the above mentioned matching strategy is based on simple SAD, we need to introduce metric learning

for better discrimination capability. Considering a recent progress by deep learning-based approaches not only in general computer vision and pattern recognition field but also in a specific field of biometrics such as gait recognition [55, 56, 69, 74–76, 78], we also adopt a deep learning-based method as a metric learning module. Particularly, because the proposed method defines a deformation for a pair of a probe and gallery images, a suitable network architecture should be a family of Siamese networks, where an input is given as a pair of images to be matched (i.e., binary classification whether the pair is generated from the same or different subject).

More specifically, we employ one of the state-of-the-art network architectures proposed in cross-view gait recognition [69]. Multiple network architectures are proposed in [69], and key difference among them lies in layers to start matching a pair of inputs. As demonstrated in [56], it is more advantageous to start matching at the bottom layer when a pair of inputs are well registered (e.g., cases of small view angle differences for cross-view gait recognition). We therefore adopt a strategy to match local feature at the bottom layer (LB), whose detailed network architecture is shown in Fig. 5. We will briefly introduce LB below and the reader may refer to the original paper [69] for more details.

In this network, paired convolutional filters W_1 and W'_1 followed by rectified linear unit (ReLU) [47] as an activation function, are at first applied to a probe and a gallery images, respectively, and they are then summed up at each spatial position and channel. Note that this simulates a sort of weighted subtraction, i.e., matching, between a probe and a gallery images. Thereafter, similarly to other CNN architectures, a triplet of normalization (cross-map normalization), spatial pooling, and convolution filters are applied sequentially twice. Via a full connection layer with dropout technique and softmax, two nodes at the output layer returns likelihoods of the same subject (node with 1) or different subjects (node with 0), respectively. The whole network is trained with the logistic regression loss. At a test phase, an output value at the node for different subject is used as a dissimilarity measure¹.

Note that, when combined with the proposed eigen FFD, the pair of inputs should be not an original probe image I_p and gallery image I_g but a morphed probe image $\mathcal{F}(I_p; \hat{\mathbf{u}})$ and a gallery image I_g .

4. Experiments

4.1. Data set

We conducted our experiments using a gait database collected by ourselves. Each subject was asked to walk a

¹This is essentially equivalent to use an output value at the node for the same subject as a similarity measure, as doing so in the original paper.

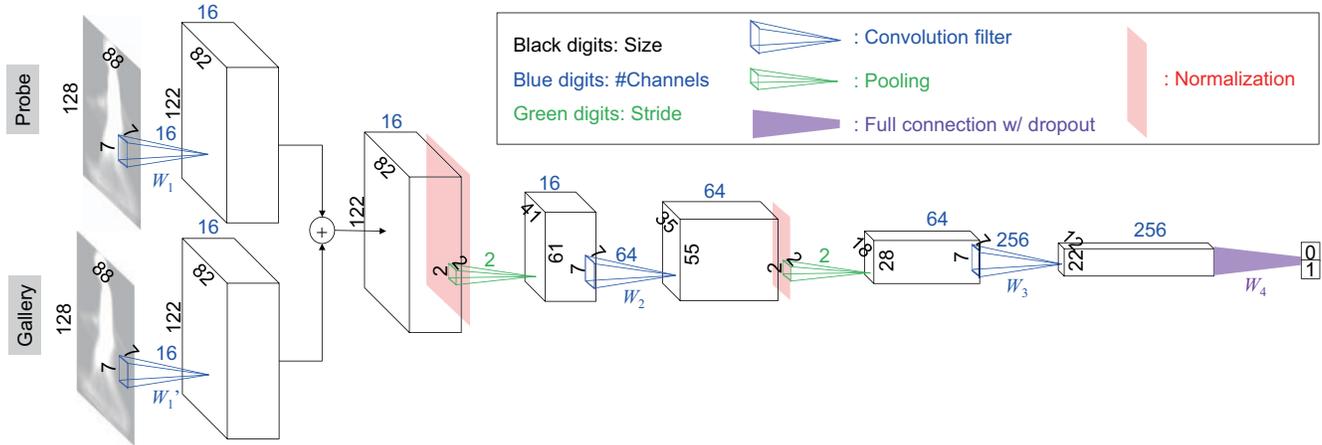


Figure 5: Network architecture of LB [69]. In this example, the size of input gait template is 88×128 pixels. Given a pair of probe and gallery images, paired convolutional filters W_1 and W'_1 with 7×7 pixels with 16 channels are applied and then summed up. A triplet of cross-map normalization, spatial pooling with 2×2 pixels with stride 2, and a convolution filter with 7×7 pixels with 64 channels are applied. Subsequently, another triplet of the same normalization, the same spatial pooling, and a convolution filter with 7×7 with 256 channels are applied. Finally, via a full connection layer with dropout technique and softmax, an output layer returns likelihoods of the same subject (node with 1) or different subjects (node with 0).

straight path surrounded by green chroma-key background twice, and was captured from a side view at approximately 8 meter distance at 5 meter height with a network camera AXIS Communications, Q1614 ($1,280 \times 960$ pixels at 25 fps). Each subject agreed with the use of captured data for the research purpose. As such, two walking image sequences of 1,334 subjects were collected, where some subjects might change his/her posture between two sequences.

The whole dataset was divided into three subsets: a training set, a gallery set, and a probe set. The training set contains 2,068 sequences of 1,034 subjects, while the gallery and probe sets form a test set composed of the rest 300 subjects that is disjoint from the 1,034 subjects in the training set. The first gait image sequences of the test set were assigned to the probe set, while the second ones were assigned to the gallery set.

As described in the introduction section, we can use any types of image-based gait templates for the proposed method. We therefore adopted GEI in our work because it is the most widely used in the gait recognition community. For this purpose, we extracted silhouette sequences from original image sequences by chroma-key and then subsequently obtained normalized silhouette sequences with 88×128 pixels by subject height normalization and registration by silhouette region centers. We then detected a gait period by maximizing auto-correlation along the temporal axis, and then average the normalized silhouette sequence over one gait period to obtain an averaged silhouette [36] a.k.a. GEI [19].

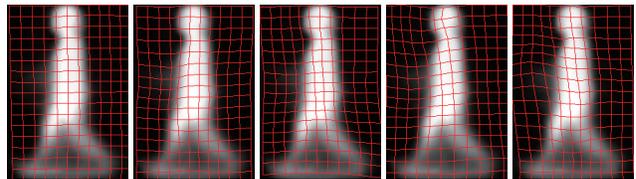


Figure 6: Eigen deformation field represented by deformed control points. From left to right, the first to the fifth eigen deformation modes are shown. For visibility, the norm of the deformation is magnified.

4.2. Parameter settings

The proposed method contains some hyper-parameters. As for the number of control points, we set $M_x = 12$ and $M_y = 17$, which sums up to $M = 204$ control points over an entire image. Moreover, we experimentally set the hyper-parameter for smoothness control in FFD as $\lambda = 100$. In addition, the dimension of the eigen FFD was decided to 5 so as that the cumulative contribution rate was over 75%.

4.3. Learnt eigen FFDs

In this subsection, we analyze the eigen FFDs learnt using the same subjects' pairs in the training set. We can observe affine-like deformations, i.e., horizontal translation and scaling in the first and the third eigen FFDs, respectively, although non-rigid deformation is partly seen in the third eigen FFD (e.g., a foot region at the backside). On the other hand, more non-rigid deformations appear in the other

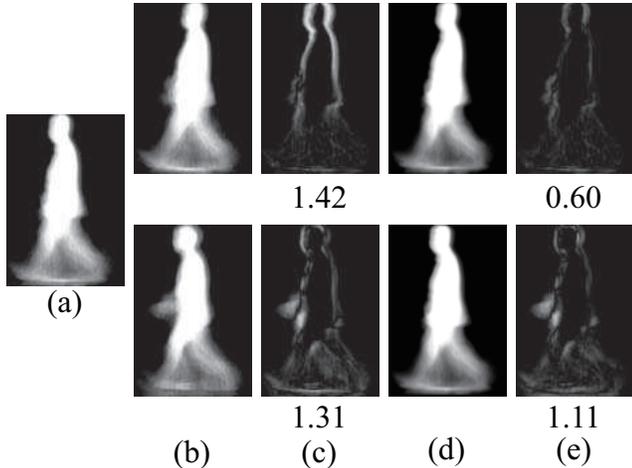


Figure 7: A successful matching example (top: the same subject, bottom: a different subject). (a) Probe I_p , (b) Gallery I_g , (c) Difference image between (a) and (b), (d) Deformed probe $F(I_p; \hat{\mathbf{u}})$ by the intra-subject deformation model, (e) Difference image between (d) and (b). A numerical value below each difference image means dissimilarity measure ($\times 10^5$).

eigen FFDs, e.g., differences in the arm swings in the third and the fifth eigen FFDs, posture change by looking down in the fourth eigen FFD, differences in the stride in the second, fourth, and the fifth eigen FFDs. These eigen FFDs encode representative intra-subject deformation modes.

4.4. Qualitative evaluation

In this subsection, we show a typical matching example without and with the learnt eigen FFD in Fig. 7 to illustrate the effectiveness of the proposed method qualitatively. A probe (Fig. 7 (a)) is matched with two galleries (Fig. 7 (b)): one from the same subject and the other from a different subject. We can see that upper-body posture of the probe is slightly different from the gallery of the same subject, and hence a dissimilarity measure for the same subject (1.42×10^5) is larger than that for the different subject (1.31×10^5), which results in false match to the different subject (Fig. 7 (c)).

On the other hands, by applying the proposed eigen FFD to the probe (Fig. 7 (d)), the above mentioned posture difference is successfully mitigated and hence the morphed probe becomes close to the gallery of the same subject, while the inter-subject difference (e.g., almost no arm swing for the probe, but some arm swings for the gallery of the different subject) is still kept to some extent. As such, the dissimilarity measure of the same subject is significantly reduced from 1.42×10^5 to 0.60×10^5 , while that for the different subject is just slightly reduced from 1.31×10^5 to 1.11×10^5

Table 1: EER [%] w/o and w/ z-normalization (denoted as EER and zEER, respectively), rank-1, rank-5, and rank-10 identification rates (denoted as Rank-1, Rank-5, and Rank-10, respectively) [%]. Bold indicates the best performances.

Method	EER	zEER	Rank-1	Rank-5	Rank-10
Direct	10.3	7.0	79.0	88.0	91.0
Raw FFD	12.3	8.5	78.0	87.3	89.0
Eigen FFD	7.3	3.7	91.0	95.3	96.7

(Fig. 7 (c) and (e)). As a result, the probe is truly matched to the gallery of the same subject.

4.5. Quantitative evaluation

In this subsection, we analyze the effect of the deformable registration without metric learning by the recognition accuracy quantitatively. We consider a direct matching without deformation (denoted as Direct) as a baseline, and also consider matching between a probe with raw deformation \mathbf{u} (denoted as Raw FFD) and a gallery as another baseline. We evaluated the proposed method (denoted as Eigen FFD) as well as the two baselines in both verification (one-to-one matching) and identification (one-to-many matching) modes.

For the verification mode, given a pair of inputs, we accept it as the same subject's pair if the dissimilarity measure between them is below an acceptance threshold, otherwise we reject it (i.e., regard it as a different subjects' pair). Here, we consider two types of error rates as performance measures: false acceptance rate (FAR) of the different subjects' pairs, and false rejection rate (FRR) of the same subjects' pairs. Because the FAR and FRR change as the acceptance threshold changes, we evaluate a trade-off between the FAR and FAR by a receiver operating characteristics (ROC) curve. In addition, we extract an equal error rate (EER) of the FAR and FRR as a typical performance measure. Moreover, we consider probe-dependent z-normalization [4] for better performance, i.e., we compute the dissimilarity scores between a specific probe and all the galleries, and normalize them so as that their means and standard deviations are 0 and 1, respectively. We report performances both without and with z-normalization.

For the identification mode, we match a probe to all the galleries and make a rank list based on dissimilarity measures (i.e., galleries with smaller dissimilarity measures get smaller ranks). We evaluated rates of true match galleries included up to rank- n by a cumulative matching characteristic (CMC) curve.

ROC curves without and with z-normalization and CMC curves are shown in Fig. 8. In addition, EERs without and with z-normalization, and rank-1, rank-5, and rank-

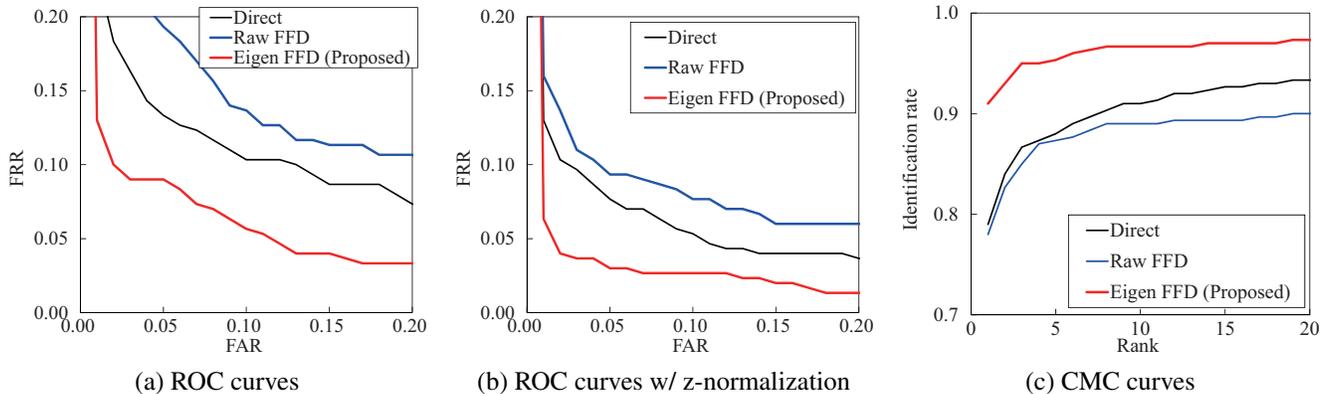


Figure 8: ROC curves w/o and w/ z-normalization and CMC curves.

Table 2: EER [%] w/o and w/ z-normalization (denoted as EER and zEER, respectively), rank-1, rank-5, and rank-10 identification rates (denoted as Rank-1, Rank-5, and Rank-10, respectively) [%] with LB. Bold indicates the best performances.

Method	EER	zEER	Rank-1	Rank-5	Rank-10
Direct	4.2	3.0	89.3	96.0	97.0
Raw FFD	7.3	6.3	66.3	87.0	92.7
Eigen FFD	3.7	2.6	92.0	98.0	98.0

10 identification rates are summarized in Table 1. We can see the proposed method yields the lowest errors and the highest identification rates among the benchmarks. More specifically, compared with direct matching without deformation, the proposed method reduces EERs without and with z-normalization by 3.0% and 3.3%, respectively, and improves rank-1, rank-5, and rank-10 identification rates by 12%, 7.3%, 5.7%, respectively. Consequently, we conclude that the proposed eigen FFD steadily improves both verification and identification performances.

4.6. Quantitative evaluation with metric learning

In this subsection, we compare the proposed method to the two benchmarks with metric learning, i.e., LB. Specifically, since the training size of this database is acceptable for learning the eigen FFD but still insufficient for training a deep neural network. Therefore, we employed the OU-ISIR gait database, the large population data set [23] with more than 4,000 subjects with four slight view angle variations for the purpose of training the deep neural network.

Similarly to the previous subsection, we evaluated the performance of the verification and identification modes by ROC curves without and with z-normalization and CMC curves as shown in Fig. 9. In addition, we summarize EERs without and with z-normalization, and rank-1, rank-5, and

rank-10 identification rates in Table 2.

Although not only the accuracy of the proposed method but also that of direct matching are improved by metric learning, the proposed method still achieves the best performance among them. Concretely speaking, the proposed method yields lower EERs without and with z-normalization than direct matching by 1.5% and 0.4%, respectively, and higher rank-1, rank-5, and rank-10 identification rates by 2.7%, 2.0%, and 1.0%, respectively. Therefore, we conclude that the proposed method is still effective even after applying the metric learning.

4.7. Limitation

Although the effectiveness of the proposed eigen FFD was demonstrated through the aforementioned experiments, there were still some failure cases where the proposed method performs poorly. We show a failure mode of matching in Fig. 10 to discuss the limitations. In this example, a gallery of the same subject’s pair moves his head forward within one gait period, and hence blurry region is found around the head in the resultant GEI (Fig. 10 (b), top), which results in a certain amount of dissimilarity. This kind of blurry region cannot be successfully mitigated by the deformable registration, and hence the dissimilarity measure is not largely reduced even after applying the eigen FFD (Fig. 10 (e), top).

On the other hand, a gallery of a different subject does not have such a blurry region at the head (Fig. 10 (b), bottom), but its stride and posture are different from a probe (Fig. 10 (a)), which results in relatively large dissimilarity measure before deformable registration (Fig. 10 (c), bottom). The differences are, however, effectively mitigated by the deformable registration and hence the dissimilarity measure is greatly reduced. As a result, the dissimilarity measure for the different subject becomes smaller than that for the same subject, and hence the probe is falsely matched to the different subject.

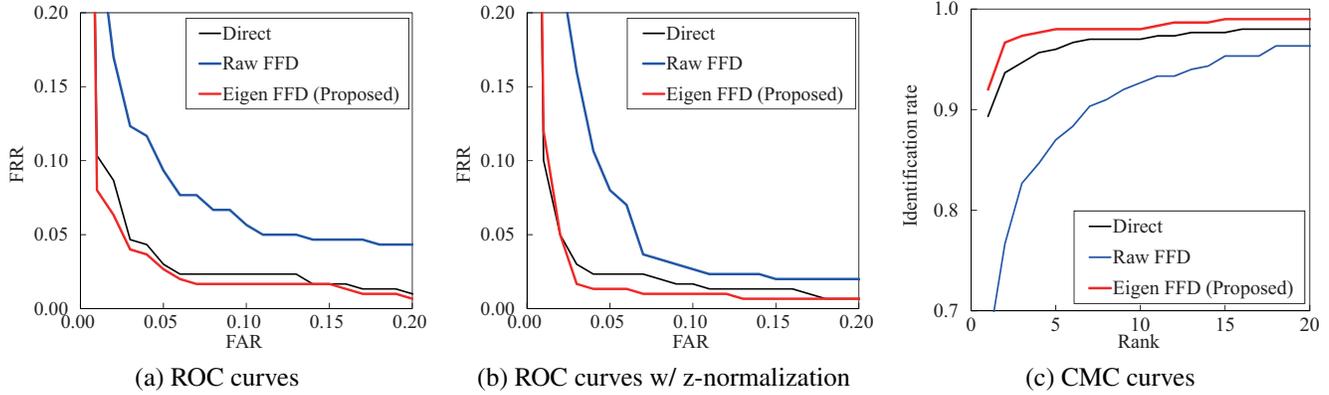


Figure 9: ROC curves w/o and w/ z-normalization and CMC curves with LB.

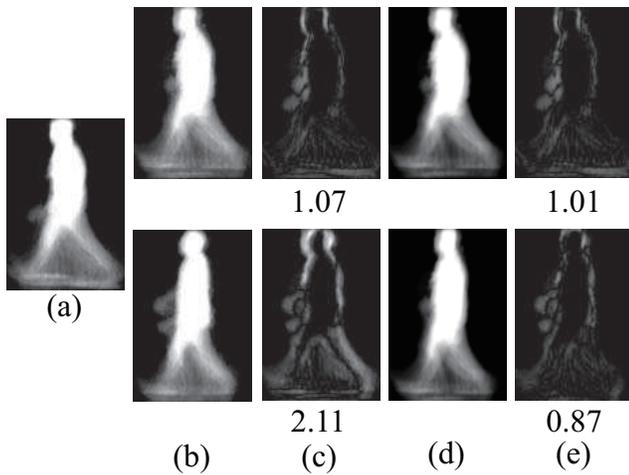


Figure 10: A failure mode (top: the same subject, bottom: a different subject). (a) Probe I_p , (b) Gallery I_g , (c) Difference image between (a) and (b), (d) Deformed probe $F(I_p; \hat{u})$ by the intra-subject deformation model, (e) Difference image between (d) and (b). A numerical value below each difference image means dissimilarity measure ($\times 10^5$).

The reasons for the failure mode are two folded. One is that the posture change within one gait period may cause a blurry region in aggregation-based gait representation such as GEI, which cannot be successfully mitigated just by applying the deformable registration to the aggregated gait feature. In order to cope with the temporal posture change within one gait period, we need to consider frame-by-frame deformable registration before aggregating over one gait period.

The other is that the eigen FFD does not necessarily contain pure intra-subject deformations even if it is extracted only from a training set of the same subjects' pairs. In fact, while stride change and posture change occur for some same subjects' pairs, such differences may also appear for

different subjects' pairs. In other words, some modes in the eigen FFD may be common both in intra-subject and inter-subject deformation. Therefore, we need to consider not only the intra-subject deformations but also the inter-subject deformations when extracting the eigen FFDs so as to keep the dissimilarity measure for a different subjects' pair large to some extent.

5. Conclusion

This paper described a method of gait recognition by deformable registration. We employed FFD with lattice-type control points and extracted the eigen FFD from a set of intra-subject deformation fields to constrain the deformation modes. Metric learning by LB, a recent deep learning framework, further improved the discrimination capability after the pre-processing deformable registration by the eigen FFD. Experiments with 1,334 subjects showed the effectiveness of the proposed method compared with direct matching without deformable registration.

One of future research avenue is consideration of not only the intra-subject deformations but also the inter-subject deformations to extract better deformation modes for discrimination. Moreover, we will consider frame-by-frame deformation to cope with posture changes within one gait period. Finally, we would like to tackle more challenging posture changes by climbing a slope or stairs, or by carrying a relatively heavy object. In addition, we will conduct experiments using more publicly available gait databases to further validate the effectiveness of the proposed method.

Acknowledgement

This work was supported by JSPS Grants-in-Aid for Scientific Research (A) JP15H01693, 18H04115, and Nanjing University of Science and Technology.

References

- [1] M. Alotaibi and A. Mahmood. Improved gait recognition based on specialized deep convolutional neural network. *Computer Vision and Image Understanding*, 164(Supplement C):103 – 110, 2017. Deep Learning for Computer Vision. 3
- [2] M. R. Aqmar, K. Shinoda, and S. Furui. Robust gait recognition against speed variation. In *Proc. of the 20th International Conference on Pattern Recognition*, pages 2190–2193, Istanbul, Turkey, Aug. 2010. 1
- [3] G. Ariyanto and M. Nixon. Marionette mass-spring model for 3d gait biometrics. In *Proc. of the 5th IAPR International Conference on Biometrics*, pages 354–359, March 2012. 2
- [4] R. Auckenthaler, M. Carey, and H. Lloyd-Thomas. Score normalization for text-independent speaker verification systems. *Digital Signal Processing*, 10(1-3):42–54, 2000. 6
- [5] K. Bashir, T. Xiang, and S. Gong. Gait recognition using gait entropy image. In *Proc. of the 3rd Int. Conf. on Imaging for Crime Detection and Prevention*, pages 1–6, Dec. 2009. 1, 2
- [6] K. Bashir, T. Xiang, and S. Gong. Gait recognition without subject cooperation. *Pattern Recognition Letters*, 31(13):2052–2060, Oct. 2010. 1
- [7] A. Bobick and A. Johnson. Gait recognition using static activity-specific parameters. In *Proc. of the 14th IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 423–430, 2001. 2
- [8] V. N. Boddeti, M. Roh, J. Shin, T. Oguri, and T. Kanade. Face alignment robust to pose, expressions and occlusions. *CoRR*, abs/1707.05938, 2017. 2
- [9] I. Bouchrika, M. Goffredo, J. Carter, and M. Nixon. On using gait in forensic biometrics. *Journal of Forensic Sciences*, 56(4):882–889, 2011. 1
- [10] F. M. Castro, M. J. Marín-Jiménez, N. Guil, and N. Pérez de la Blanca. Automatic learning of gait signatures for people identification. In I. Rojas, G. Joya, and A. Catala, editors, *Advances in Computational Intelligence: 14th International Work-Conference on Artificial Neural Networks, IWANN 2017, Cadiz, Spain, June 14-16, 2017, Proceedings, Part II*, pages 257–270, Cham, 2017. Springer International Publishing. 3
- [11] C. Chen, J. Liang, H. Zhao, H. Hu, and J. Tian. Frame difference energy image for gait recognition with incomplete silhouettes. *Pattern Recognition Letters*, 30(11):977 – 984, 2009. 2
- [12] T. Cootes, C. Taylor, D. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995. 2
- [13] D. Cunado, M. Nixon, and J. Carter. Automatic extraction and description of human gait models for recognition purposes. *Computer Vision and Image Understanding*, 90(1):1–41, 2003. 2
- [14] B. Decann and A. Ross. Gait curves for human recognition, backpack detection, and silhouette correction in a nighttime environment. In *Proc. of the SPIE, Biometric Technology for Human Identification VII*, volume 7667, pages 1–13, 2010. 1
- [15] 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 *Proc. of the 4th Asian Conf. on Pattern Recognition (ACPR 2017)*, pages 929–934, Nov. 2017. 2
- [16] 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. 1
- [17] Y. Guan, C. T. Li, and Y. Hu. Robust clothing-invariant gait recognition. In *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2012 Eighth International Conference on*, pages 321–324, July 2012. 2, 3
- [18] Y. Guan, C. T. 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. 2, 3
- [19] J. Han and B. Bhanu. Individual recognition using gait energy image. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 28(2):316–322, 2006. 1, 2, 3, 5
- [20] V. N. T. Hoang Thai. Face alignment using active shape model and support vector machine. *International Journal of Biometric and Bioinformatics*, 4(6):224–234, 2011. 2
- [21] M. A. Hossain, Y. Makihara, J. Wang, and Y. Yagi. Clothing-invariant gait identification using part-based clothing categorization and adaptive weight control. *Pattern Recognition*, 43(6):2281–2291, Jun. 2010. 1
- [22] H. Iwama, D. Muramatsu, Y. Makihara, and Y. Yagi. Gait verification system for criminal investigation. *IPSJ Trans. on Computer Vision and Applications*, 5:163–175, Oct. 2013. 1
- [23] H. Iwama, M. Okumura, Y. Makihara, and Y. Yagi. The ou-isir gait database comprising the large population dataset and performance evaluation of gait recognition. *IEEE Transactions on Information Forensics and Security*, 7(5):1511–1521, Oct. 2012. 7
- [24] Z. Ji, W. Zheng, N. Sun, C. Zou, and L. Zhao. Face alignment using an improved active shape model. In D.-S. Huang, K. Li, and G. W. Irwin, editors, *International Conference on Intelligent Computing (ICIC 2006)*, pages 846–851, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg. 2
- [25] A. Jourabloo, M. Ye, X. Liu, and L. Ren. Pose-invariant face alignment with a single cnn. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 3219–3228, Oct 2017. 2
- [26] A. Kale, A. Roy-Chowdhury, and R. Chellappa. Fusion of gait and face for human identification. In *Proc. of the IEEE Int. Conf. on Acoustics, Speech, and Signal Processing 2004 (ICASSP'04)*, volume 5, pages 901–904, 2004. 1
- [27] W. Kusakunniran. Attribute-based learning for gait recognition using spatio-temporal interest points. *Image Vision Comput.*, 32(12):1117–1126, Dec. 2014. 1
- [28] W. Kusakunniran, Q. Wu, J. Zhang, and H. Li. Support vector regression for multi-view gait recognition based on local motion feature selection. In *Proc. of IEEE computer society conference on Computer Vision and Pattern Recognition 2010*, pages 1–8, San Francisco, CA, USA, Jun. 2010. 1
- [29] 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. 1
- [30] W. Kusakunniran, Q. Wu, J. Zhang, and H. Li. Cross-view and multi-view gait recognitions based on view transformation model using multi-layer perceptron. *Pattern Recognition Letters*, 33(7):882–889, 2012. 1
- [31] 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. 1, 2
- [32] 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. 1
- [33] T. H. W. Lam, K. H. Cheung, and J. N. K. Liu. Gait flow image: A silhouette-based gait representation for human identification. *Pattern Recognition*, 44:973–987, April 2011. 2
- [34] L. Lee. *Gait Analysis for Classification*. PhD thesis, Massachusetts Institute of Technology, 2002. 2
- [35] 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. 1
- [36] Z. Liu and S. Sarkar. Simplest representation yet for gait recognition: Averaged silhouette. In *Proc. of the 17th International Conference on Pattern Recognition*, volume 1, pages 211–214, Aug. 2004. 5
- [37] Z. Liu and S. Sarkar. Effect of silhouette quality on hard problems in gait recognition. *IEEE Trans. of Systems, Man, and Cybernetics Part B: Cybernetics*, 35(2):170–183, 2005. 2, 3
- [38] Z. Liu and S. Sarkar. Improved gait recognition by gait dynamics normalization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(6):863–876, 2006. 1
- [39] J. Lu and Y.-P. Tan. Uncorrelated discriminant simplex analysis for view-invariant gait signal computing. *Pattern Recognition Letters*, 31(5):382–393, 2010. 1
- [40] N. Lynnerup and P. Larsen. Gait as evidence. *IET Biometrics*, 3(2):47–54, 6 2014. 1
- [41] 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. 1, 2
- [42] 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. 1, 2
- [43] A. Mansur, Y. Makihara, R. Aqmar, and Y. Yagi. Gait recognition under speed transition. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 2521–2528, June 2014. 1
- [44] R. Martin-Felez and T. Xiang. Uncooperative gait recognition by learning to rank. *Pattern Recognition*, 47(12):3793 – 3806, 2014. 1
- [45] A. Mori, Y. Makihara, and Y. Yagi. Gait recognition using period-based phase synchronization for low frame-rate videos. In *Proc. of the 20th International Conference on Pattern Recognition*, pages 2194–2197, Istanbul, Turkey, Aug. 2010. 1
- [46] H. Murase and R. Sakai. Moving object recognition in eigenspace representation: Gait analysis and lip reading. *Pattern Recognition Letters*, 17:155–162, 1996. 1
- [47] V. Nair and G. E. Hinton. Rectified linear units improve restricted boltzmann machines. In J. Furnkranz and T. Joachims, editors, *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 807–814. Omnipress, 2010. 4
- [48] M. S. Nixon, T. N. Tan, and R. Chellappa. *Human Identification Based on Gait*. Int. Series on Biometrics. Springer-Verlag, Dec. 2005. 1
- [49] R. S. D. S. Z. S. of Engineering. Facial expression analysis using active shape model. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 8(1):9–22, 2015. 2
- [50] M. Rokanujjaman, M. Islam, M. Hossain, M. Islam, Y. Makihara, and Y. Yagi. Effective part-based gait identification using frequency-domain gait entropy features. *Multimedia Tools and Applications*, 74(9):3099–3120, May 2015. 1
- [51] A. Ruta, Y. Li, and X. Liu. Low-resolution human detection and gait recognition in natural scenes. In *Prof. of the 14th Int. Conf. on Automation and Computing*, pages 1–6, Sep. 2008. 1
- [52] S. Sarkar, J. Phillips, Z. Liu, I. Vega, P. G. ther, and K. Bowyer. The humanid gait challenge problem: Data sets, performance, and analysis. *IEEE Trans. of Pattern Analysis and Machine Intelligence*, 27(2):162–177, 2005. 1
- [53] T. W. Sederberg and S. R. Parry. Free-form deformation of solid geometric models. *ACM SIGGRAPH computer graphics*, 20(4):151–160, 1986. 2, 3
- [54] G. Shakhnarovich and T. Darrell. On probabilistic combination of face and gait cues for identification. In *Proc. Automatic Face and Gesture Recognition 2002*, volume 5, pages 169–174, 2002. 1
- [55] K. Shiraga, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi. Geinet: View-invariant gait recognition using a convolutional neural network. In *Proc. of the 8th IAPR Int. Conf. on Biometrics (ICB 2016)*, number O19, pages 1–8, Halmstad, Sweden, Jun. 2016. 1, 4
- [56] 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*, PP(99):1–1, 2017. 4
- [57] R. Tanawongsuwan and A. Bobick. Gait recognition from time-normalized joint-angle trajectories in the walking plane. In *Proc. of the 14th IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 726–731, Jun. 2001. 1
- [58] R. Tanawongsuwan and A. Bobick. A study of human gaits across different speeds. Technical report, Georgia Tech, 2003. 2

- [59] 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. 1, 2, 3
- [60] D. Tao, X. Li, X. Wu, and S. J. Maybank. General tensor discriminant analysis and gabor features for gait recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(10):1700–1715, Oct. 2007. 1
- [61] 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. 2, 3
- [62] N. F. Troje. Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of Vision*, 2:371–387, 2002. 2
- [63] R. Urtasun and P. Fua. 3d tracking for gait characterization and recognition. In *Proc. of the 6th IEEE International Conference on Automatic Face and Gesture Recognition*, pages 17–22, 2004. 2
- [64] D. Wagg and M. Nixon. On automated model-based extraction and analysis of gait. In *Proc. of the 6th IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pages 11–16, 2004. 2
- [65] C. Wang, J. Zhang, L. Wang, J. Pu, and X. Yuan. Human identification using temporal information preserving gait template. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 34(11):2164–2176, nov. 2012. 1, 2, 3
- [66] L. Wang, T. Tan, H. Ning, and W. Hu. Silhouette analysis-based gait recognition for human identification. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(12):1505–1518, dec. 2003. 1
- [67] T. Wolf, M. Babaei, and G. Rigoll. Multi-view gait recognition using 3d convolutional neural networks. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 4165–4169, Sept 2016. 3
- [68] 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. 3
- [69] 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*, PP(99):1–1, 2016. 1, 2, 3, 4, 5
- [70] P. Xiong, L. Huang, and C. Liu. Initialization and pose alignment in active shape model. In *Proc. of 20th International Conference on Pattern Recognition*, pages 3971–3974, Aug 2010. 2
- [71] 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. 2, 3
- [72] C. Yam, M. Nixon, and J. Carter. Automated person recognition by walking and running via model-based approaches. *Pattern Recognition*, 37(5):1057–1072, 2004. 2
- [73] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, and H.-J. Zhang. Discriminant analysis with tensor representation. In *Proc. of the IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, pages 526–532, Jun. 2005. 2, 3
- [74] 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. 3, 4
- [75] S. Yu, H. Chen, Q. Wang, L. Shen, and Y. Huang. Invariant feature extraction for gait recognition using only one uniform model. *Neurocomput.*, 239(C):81–93, May 2017. 3, 4
- [76] 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, March 2016. 3, 4
- [77] J. Zhang, J. Pu, C. Chen, and R. Fleischer. Low-resolution gait recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 40(4):986–996, Aug 2010. 1
- [78] Z. Zhang, J. Chen, Q. Wu, and L. Shao. Gii representation-based cross-view gait recognition by discriminative projection with list-wise constraints. *IEEE Transactions on Cybernetics*, PP(99):1–13, 2017. 3, 4
- [79] G. Zhao, G. Liu, H. Li, and M. Pietikainen. 3d gait recognition using multiple cameras. In *7th International Conference on Automatic Face and Gesture Recognition (FG06)*, pages 529–534, April 2006. 2