applied sciences

MDPI

Article

Gait Energy Response Functions for Gait Recognition against Various Clothing and Carrying Status

Xiang Li [1,2,*], Yasushi Makihara [2], Chi Xu [1,2], Daigo Muramatsu [2], Yasushi Yagi [2] and Mingwu Ren [1]

1 School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China; xuchisherry@gmail.com (C.X.); renmingwu@mail.njust.edu.cn (M.R.)
2 The Institute of Scientific and Industrial Research, Osaka University, Osaka 567-0046, Japan; makihara@am.sanken.osaka-u.ac.jp (Y.M.); muramatsu@am.sanken.osaka-u.ac.jp (D.M.); yagi@am.sanken.osaka-u.ac.jp (Y.Y.)
* Correspondence: lixiangmzlx@gmail.com; Tel.: +86-138-0516-3641

Received: 24 July 2018; Accepted: 13 August 2018; Published: 16 August 2018

Abstract: Silhouette-based gait representations are widely used in the current gait recognition community due to their effectiveness and efficiency, but they are subject to changes in covariate conditions such as clothing and carrying status. Therefore, we propose a gait energy response function (GERF) that transforms a gait energy (i.e., an intensity value) of a silhouette-based gait feature into a value more suitable for handling these covariate conditions. Additionally, since the discrimination capability of gait energies, as well as the degree to which they are affected by the covariate conditions, differs among body parts, we extend the GERF framework to spatially dependent GERF (SD-GERF) which accounts for spatial dependence. Moreover, the proposed GERFs are represented as a vector in the transformation lookup table and are optimized through an efficient generalized eigenvalue problem in a closed form. Finally, two post-processing techniques, Gabor filtering and spatial metric learning, are employed for the transformed gait features to boost the accuracy. Experimental results with three publicly available datasets including clothing and carrying status variations show the state-of-the-art performance of the proposed method compared with other state-of-the-art methods.

Keywords: gait recognition; gait energy response function; spatial dependence; Gabor filtering; metric learning

1. Introduction

Gait, as a behavioral biometric, has its own superior property to other biometrics (e.g., iris, face, finger veins) for person recognition, i.e., it can be used at a long distance by a camera with low image resolution. Additionally, it can be regarded as an unconscious behavior because people usually never conceal their gait deliberately. Therefore, gait recognition [1] is a promising key technology for many real-world applications such as surveillance, forensics, and criminal investigation [2–4].

Approaches to gait recognition are mainly separated into two families: model-based [5–8] and appearance-based [9–13]. The former one usually fits a human model to an input image at first and then extracts both motion information (e.g., joint angle sequences) and static information (e.g., body shapes) as gait features, while the latter directly extracts gait features from input images (silhouette images in many cases) without model fitting. Thus, appearance-based approaches are more feasible in real applications, which can still be applied to low-resolution videos, when model-based approaches are difficult to fit the human model correctly.

In the literature, appearance-based gait representations mainly include motion-based features [14,15] and silhouette-based features, where the latter one has often been used for gait recognition because of their simple yet effective properties, such as gait energy image (GEI) [11], frequency-domain feature

(FDF) [16], chrono-gait image [17], and Gabor GEI [18]. Among them, GEI, a.k.a. averaged silhouette [19] (see Figure 1, top row, as an example) is the most frequently used gait feature because it can be generated easily by averaging the silhouettes over a gait period, which makes it relatively robust to segmentation errors. The GEI also effectively represents both static and dynamic components with a single template (e.g., gait energies: intensity values 0 and 255 represent static components of background and foreground respectively, while intermediate grayscale values such as 127 represent dynamic components). However, these appearance-based gait representations of individuals are easily changed by various covariates (e.g., clothing and carrying status), which induces a serious decline in recognition accuracy.

To maintain robustness of gait recognition against these covariates, existing appearance-based methods fall into two main families. The first one is spatial metric learning-based approaches, which concentrate on learning a more discriminant feature space from original appearance-based features to achieve better performance against the covariates. The second one is intensity transformation-based approaches, which more care about feature representation aspects. Specifically, the intensity transformation-based approaches transform intensity values of an original gait feature (e.g., gait energies in the case of GEI) into more discriminative values to increase the robustness against change of the covariate conditions.

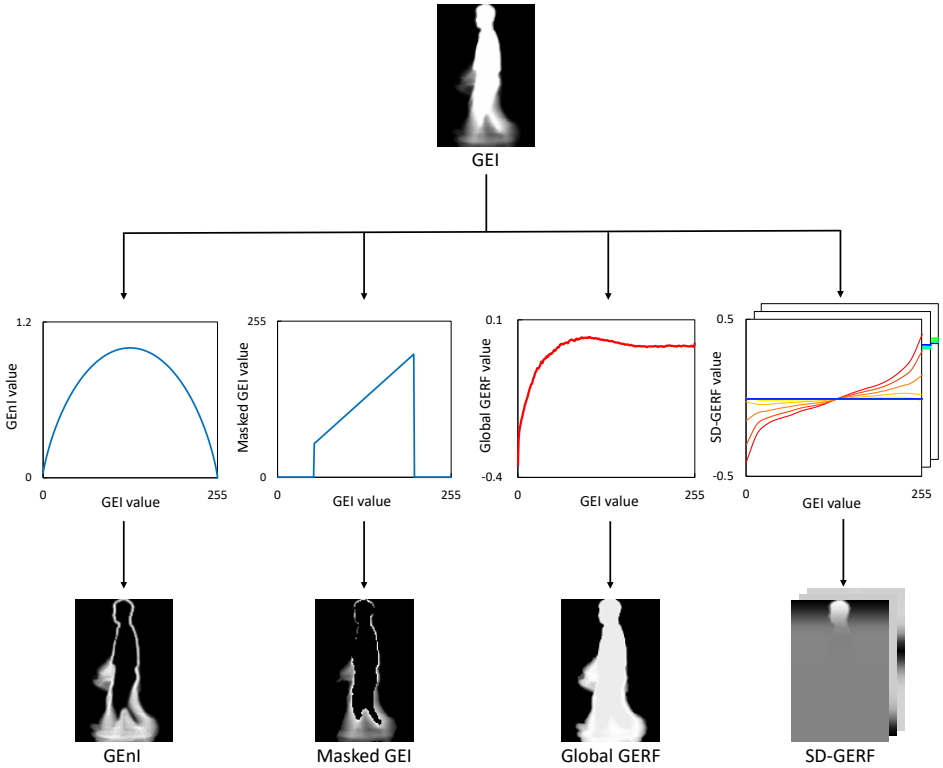

Figure 1. Concept of the proposed gait energy response functions (GERFs) as well as other existing intensity transformation-based approaches. In the fourth column, multiple profiles denote spatially dependent response functions for individual vertical positions ranging from upper position (red line) to lower position (blue line), and multiple images denote a set of spatially dependent response functions.

A gait entropy image (GEnI) [12] is a typical intensity transformation-based gait feature whose intensities are transformed from the GEI. Because the gait energy of each pixel is regarded as a foreground probability in the GEI representation, Shannon entropy for the foreground probability at each pixel is computed as a transformed intensity at each pixel in the GEnI, to enhance the dynamic components while attenuating the static components. For example, the intensities of the pixels with small and large grayscale values (e.g., 0 and 255) in the GEI become small in the GEnI, while the intensities of the pixels with middle grayscale values (e.g., 127) in the GEI become large (see Figure 1,

first column). This means that the dynamic components are enhanced while the static components are attenuated, regardless of whether they come from a static foreground or background in the GEnI representation. Therefore, the GEnI is insensitive to the static component changes derived from clothing and carrying status variations to some extent. However, the static components discarded in the GEnI representation, still have some discrimination capability even under variations in clothing and carrying status. Moreover, another negative aspect of the GEnI is that a pair of different gait energies that are symmetrical with respect to the intermediate value (i.e., $127.5 - x$ and $127.5 + x$), are transformed into the same intensity value in the GEnI (see Figure 1, first column); i.e., the dynamic component is also sacrificed.

Masked GEI [13], as another intensity transformation-based gait feature, has been proposed to address the latter problem. It is generated by masking out (i.e., setting to zero) those gait energies with smaller gait entropy than a threshold, while the others are kept unchanged (see Figure 1, second column). Masked GEI therefore keeps the dynamic components as they are, while discarding useful static components. In addition, the threshold for masking out is a sensitive parameter and hence requires careful treatment.

Based on the previously described observation that both GEnI and masked GEI are generated from GEI, we employ a mapping function named gait energy response function (GERF) to describe these gait energy transformation processes. Both use handcrafted GERFs, which more care about the dynamic components. However, such simply handcrafted GERFs cannot well handle various covariates, since different covariate conditions affect different components of the original gait features. For example, clothing and carrying status variations affect static components such as torso and limb shapes more than dynamic components such as leg and arm motion, while the speed variation exerts a greater effect on dynamic components. Therefore, a key to the success of the gait energy response function is their appropriate enhancement and/or attenuation of the static and dynamic components in the original gait features by taking the covariate conditions into account.

We therefore proposed a more general and data-driven framework for designing the GERF to transform the GEI into more discriminative features (see Figure 1, third column) in our previous paper [20] and seek to show the effectiveness of the proposed method in gait recognition under variations in clothing and carrying status.

Furthermore, we note that the importance (i.e., discrimination capability) of static and dynamic components, as well as the degree to which they are affected by the covariate conditions, differs among the body parts. For example, the head mainly contains static components and is seldom affected by carried objects, and hence the static components may be more important than the dynamic ones. In contrast, the leg contains more dynamic components and is often affected by clothing and carried objects; therefore, the dynamic components may be more important than the static ones.

We therefore extend a spatially dependent version of GERF (called SD-GERF), and contrast it with the previous spatially independent version, which we denote as Global GERF, to consider the previously highlighted differences in the importance of static and dynamic components among the body parts. More specifically, instead of designing a single common GERF over the whole gait feature, we design multiple GERFs for individual vertical positions in the GEI (see the multiple profiles in the last column of Figure 1), and since responses for each GERF often localize to a certain body part, we further adopt a set of the spatially dependent GERFs (i.e., multiple SD-GERFs, corresponding to multiple images in the last column of Figure 1) to cover the whole body information and fuse the results of them for better accuracy. More explanations are given in Section 3.4. We summarize the contributions of this paper as follows:

(1) Data-driven approach to intensity transformation.

The proposed method learns the Global GERF in a data-driven way, unlike existing intensity transformation-based methods such as GEnI and masked GEI use handcrafted designs. Specifically, we use the training set including variation to train the Global GERF for the whole GEI to maximize the discrimination capability. This enables us to realize a good tradeoff between static and dynamic components, while existing methods only enhance the dynamic components.

(2) Extension of the Global GERF into a spatially dependent function.

We propose a spatially dependent framework for the Global GERF to consider the differences in the importance of static and dynamic components among different body parts. Specifically, we prepare individual GERFs for vertical positions of GEI as SD-GERF and train the SD-GERF simultaneously. Moreover, we exploit multiple SD-GERFs as shown in Figure 1, 4th column, and integrate their scores in a score-level fusion framework to improve accuracy.

(3) Closed-form solution for optimization.

We train to maximize the ratios of dissimilarity between different-subject pairs and same-subject pairs, and consequently formulate this optimization process as a generalized eigenvalue problem both for Global GERF and SD-GREF. We therefore obtain an analytic solution in a closed form without any iteration, and hence avoid troublesome convergence problems, which are inseparable from a nonlinear optimization framework.

(4) State-of-the-art accuracy for gait recognition under variations in clothing and carrying status.

We achieve the state-of-the-art accuracies of gait recognition under variations in clothing and carrying status on three publicly available gait databases: the OU-ISIR Gait Database, the Treadmill Dataset B [21] (OU-TD-B), the OU-ISIR Gait Database, Large Population dataset with bag β version [22] (OU-LP-Bag β), and the CASIA Gait Database B [23] (CASIA-B).

This paper is an extended version of a conference paper [20]. More specifically, the extensions include two main aspects: (1) we extend original Global GERF to SD-GERF considering the differences in the importance of static and dynamic components among different body parts; (2) we evaluate the proposed method on two other gait databases (OU-LP-Bag β and CASIA-B) containing another variation (i.e., carrying status). Experimental results show the SD-GERF achieves the state-of-the-art performance and outperforms Global GERF by a large margin in all databases.

2. Related Work

2.1. Spatial Metric Learning-Based Approaches to Gait Recognition

The spatial metric learning-based approaches concentrate on learning a more discriminant feature space from original appearance-based features to achieve better performance against the covariates. Additionally, there are two further categories within the spatial metric learning-based approaches: whole-based [11–13,24–27] and part-based approaches [28–31]. For the whole-based approaches, the holistic appearance-based features are projected into a discriminative space to make them more robust against the covariate conditions. For example, Han et al. [11] applied linear discriminant analysis (LDA) to real and synthesized templates of GEI to reduce intra-class variations (e.g., clothing variations) to some extent. Xu et al. [25] proposed a matrix representation-based subspace learning algorithm by performing a two-stage scheme composed of concurrent subspace analysis (CSA) to reduce dimension and discriminant analysis with tensor representation (DATER) to obtain discriminant subspace. A random subspace method (RSM) framework that combines multiple inductive biases also has been proposed in [26,27].

While the previously described approaches rely on whole-based representation, some studies decompose the holistic appearance-based features into multiple body part-dependent features and

enhance the effective parts for recognition while attenuating the parts affected by the covariate conditions. This is because variations such as clothing and carrying status usually affect not the whole but certain parts, and a decline in accuracy is derived mainly from the affected parts. Thus, the part-based approaches have possibilities for achieving better accuracy by appropriate treatment of the affected body parts (e.g., reducing the weights of the affected body parts for recognition). For example, Hossain et al. [28] divided the human body into eight sections based on anatomical knowledge and mitigated the effect of clothing variations by adaptively assigning larger and smaller weights to affected and unaffected sections, respectively. Iwashita et al. [31] divided the human body into several areas equally and then estimated a comparison weight for each area. Weights were based on the similarity between extracted features and those in the database for standard clothing. Rokanujjaman et al. [29] defined more effective and less effective body parts by analyzing cumulative row-wise recognition rates. The frequency domain-based gait entropy features (EnDFT) of the more effective parts were used for recognition.

2.2. Intensity Transformation-Based Approaches to Gait Recognition

As mentioned in Section 1, the intensity transformation-based approaches transform intensity values of an original gait feature into more discriminative values to increase the robustness against change of the covariate conditions. For example, Bashir et al. [12] computed the GEnI by Shannon entropy of the foreground probability at each pixel (i.e., gait energy in the GEI). The GEnI encodes the randomness of pixel values in the silhouette images over a complete gait cycle, thereby capturing more motion information (dynamic components) rather than static information, which improves robustness against shape changes (e.g., clothing and carrying status). Masked GEI [13] is another intensity transformation-based approach that keeps the dynamic components as their original values, while it zero-pads the static components (i.e., both almost foreground and background parts), which are decided by a certain threshold. In contrast to upper two approaches, recently Makihara et al. [22] proposed a joint intensity transformation-based method, which focused on the joint intensity transformation of a pair images instead of a single one. Specifically, a metric on joint intensities between a pair of probe and gallery was learned to mitigate the large intra-subject differences as well as leverage the subtle inter-subject differences. However, similarly to our previous work [20], it does not consider the different effect of the joint intensity metric among different body parts (e.g., more leverage on the motion difference in the legs than the torso part since clothing and carrying status less affect the legs).

2.3. CNN-Based Approaches to Gait Recognition

Recently, more studies on CNN-based gait recognition have been published [15,32–37]. For example, Wu et al. [32] used every raw silhouette from each gait sequence as an individual input in their network. While Wolf et al. [33] regarded raw silhouettes from each gait sequence as a spatiotemporal input and designed a 3D CNN model. Instead of using raw sequences, Shiraga et al. [34] designed an eight-layered CNN network called GEINet using averaged silhouettes (i.e., GEI). All the aforementioned networks regard gait recognition as person classification from the same gait class. Besides, unlike the GEINet which uses only one input GEI, Zhang et al. [35] and Wu et al. [36] designed their networks with two input GEIs (a pair of probe and gallery GEIs). Both of their networks try to perform similarity learning between a probe GEI and a gallery GEI, then tell whether these two GEIs come from the same person or not. In addition to silhouette-based feature GEI, motion features (e.g., optical flow maps) are also used in some approaches [15,37]. All these approaches achieved significant improvements compared with traditional methods. However, they all require massive training set to ensure their best performance.

3. Gait Recognition Using GERF

The pipeline of our proposed method is shown in Figure 2. Given two raw sequences of a gallery and probe, first extract gait silhouettes using a background subtraction-based graph-cut

segmentation [38] and obtain registered and size-normalized silhouettes using the region center information [16]. Second, average the silhouettes over a gait period to get a GEI. Third, transform the GEI using learned GERF from the training set with covariates, e.g., clothing or carrying status variations. Then, after two post-processing techniques, i.e., Gabor filtering and spatial metric learning, calculate the L2 distance as a dissimilarity score between the gallery and probe. Finally, do the comparison by comparing the score with an acceptance threshold for verification (one-to-one comparison) scenarios, or by using the most commonly used nearest neighbor classifier for identification (one-to-many comparison) scenarios.

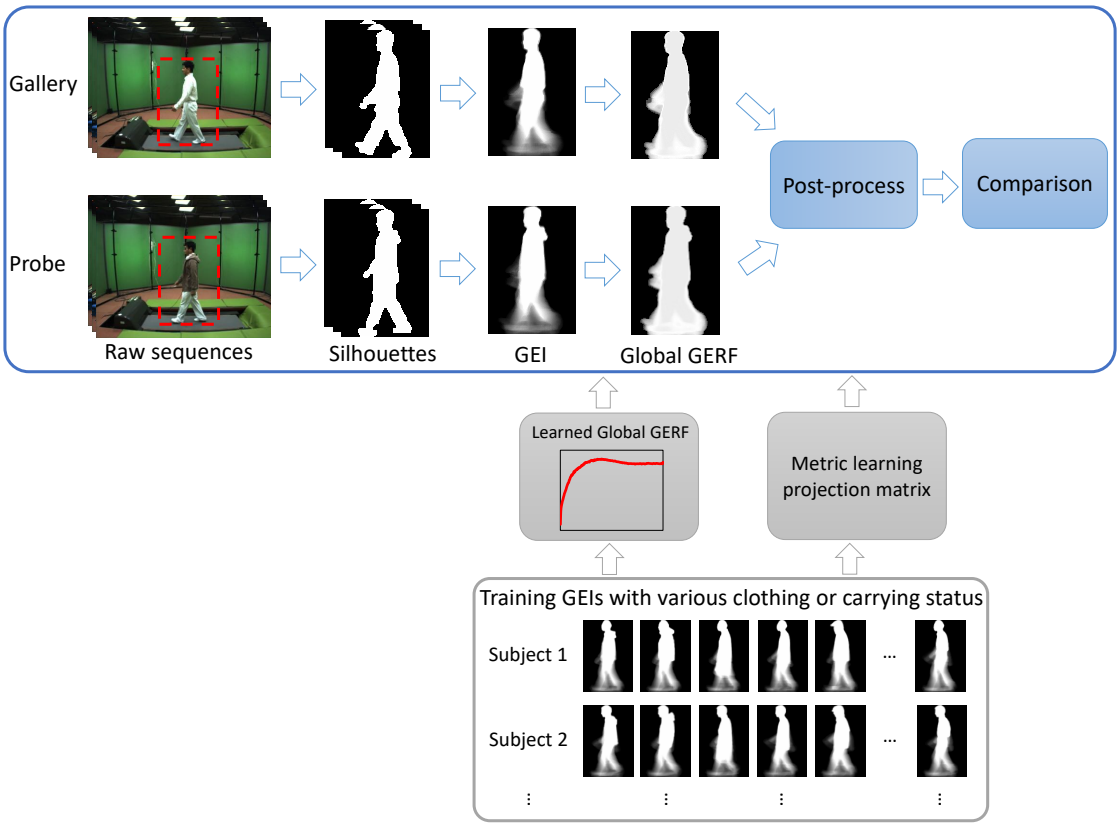

Figure 2. Pipeline of the proposed method. The learned GERF from the training set can be replaced by the extended spatially dependent GERF (SD-GERF). Here just take Global GERF as an example.

3.1. Representation of Global GERF

We apply the Global GERF for the most widely employed appearance-based gait representation, i.e., GEI. Therefore, we first introduce the concept of GEI. The GEI [11] is a gait template generated by averaging registered and size-normalized silhouettes within a complete gait period T as

$$I(x,y) = \frac{1}{T}\sum_{t=1}^{T} B(x,y,t), \tag{1}$$

where $B(x,y,t)$ is a registered and size-normalized binary silhouette value at the position (x,y) in the t-th frame (with binary value 0 and $I_{\max}$ ($I_{\max}$ is usually 255 for an image with 8-bit depth.) for background and foreground, respectively), and $I(x,y)$ is a gait energy at the same position. We approximate $I(x,y)$ as an integer for simplicity, i.e., $I(x,y) \in \{0,1,\ldots,I_{\max}\}$, since the domain of gait energy is real numbers.

The Global GERF f is defined as a transformation from an original gait energy $I(x,y)$ to a transformed one $I'(x,y)$ as follows

$$I'(x,y) = f(I(x,y)) \ \forall (x,y). \tag{2}$$

Because the original gait energy takes one of $(I_{\max}+1)$ integers from 0 to $I_{\max}$, the Global GERF is represented as a lookup table $\boldsymbol{f} = [f_0, ..., f_{I_{\max}}]^T \in \mathbb{R}^{I_{\max}+1}$, where f_i represents a transformed gait energy from an original gait energy i.

We then define a dissimilarity measure between a pair of transformed GEIs from original GEIs I_1 and I_2. We simply adopt the Euclidean distance between them and define its squared distance $d^2_{I_1,I_2}$ and further formulate it in the quadratic form of $\boldsymbol{f}$ as

$$d^2_{I_1,I_2} = \sum_{x,y} (f_{I_1(x,y)} - f_{I_2(x,y)})^2 = \boldsymbol{f}^T A_{I_1,I_2} \boldsymbol{f} \ , \tag{3}$$

where $A_{I_1,I_2} \in \mathbb{R}^{(I_{\max}+1)\times(I_{\max}+1)}$ is a coefficient matrix for quadratic-form representation and its (l,m) component can be calculated as

$$\begin{aligned} (A_{I_1,I_2})_{l,m} = \sum_{x,y} (&\delta_{I_1(x,y),l}\delta_{I_1(x,y),m} + \delta_{I_2(x,y),l}\delta_{I_2(x,y),m} \\ &- \delta_{I_1(x,y),l}\delta_{I_2(x,y),m} - \delta_{I_2(x,y),l}\delta_{I_1(x,y),m}) \ , \end{aligned} \tag{4}$$

where $\delta_{i,j}$ is the Kronecker delta defined as

$$\delta_{i,j} = \begin{cases} 1 & (i=j) \\ 0 & (i \neq j). \end{cases} \tag{5}$$

3.2. Representation of SD-GERF

Because the Global GERF trains a single common response function for the whole GEI, it cannot consider differences in the importance of static and dynamic components among the body parts (e.g., the static components are more important for the head, while the dynamic components are more important for the leg). We therefore introduce spatial dependency into the GERF framework.

The most straightforward way to do this is to define different GERFs for individual spatial positions, i.e., for every pixel, and to optimize an SD-GERF that is a concatenated vector of the multiple GERFs. This strategy, however, significantly increases the number of variables (i.e., the dimension of the SD-GREF vector in proportion to the image size) and hence suffers from generalization errors as well as an increase in space and time complexity. We therefore consider reducing the dimension of the SD-GERF vector.

First, as the body parts are divided by vertical positions in most part-based approaches to gait recognition [28,29], we also consider only vertical spatial dependency while neglecting horizontal spatial dependency. Second, because GERFs of adjacent vertical positions should be similar, in other words, abrupt changes of the GERFs among adjacent vertical positions are unlikely to occur, we set the SD-GERFs of a smaller number of vertical positions, using an interval coarser than a single pixel. We then represent in-between GERFs by linear interpolation. Third, while the Global GERF is defined by $(I_{\max}+1)$ components, i.e., at every intensity level, we represent a GERF with a smaller number of intensity levels at a coarser intensity interval, and then represent in-between response values by linear interpolation as well.

In summary, we introduce a set of control points distributed over vertical positions and intensity levels at certain intervals as shown in Figure 3 where response values are defined and estimate intermediate response values by bilinear interpolation from the adjacent control points. Note that

not only pixels on the control points but also those in-between control points are used for training and testing.

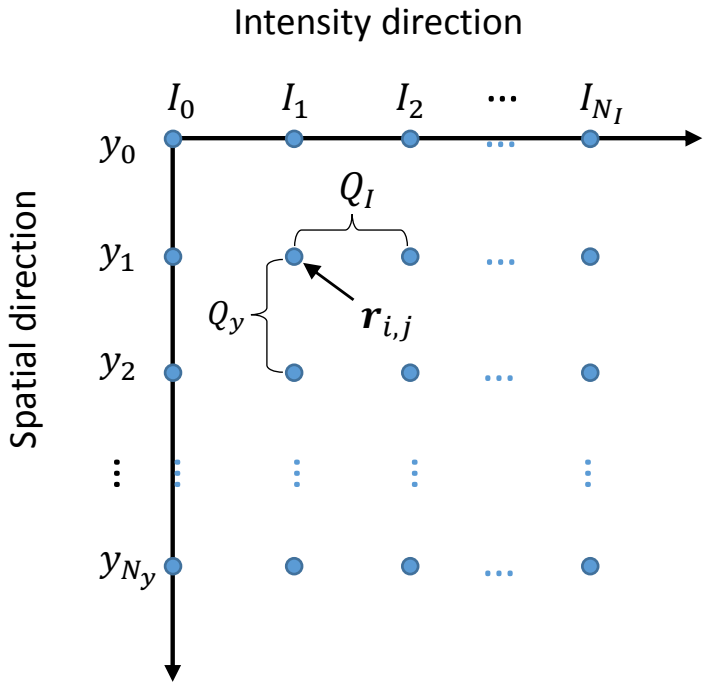

Figure 3. Representation of control points. For example, if $Q_y = 8$, then 0th, 8th, ..., rows of a gait energy image (GEI) image are chosen as vertical positions; if $Q_I = 16$, then intensities (gray values: 0, 16, ...) are chosen as intensity positions.

Next, we introduce a mathematical definition of the SD-GERF based on the previously described concept. A control point for the i-th intensity I_i and the j-th row (vertical position) y_j is defined as

$$\boldsymbol{r}_{i,j} = \left[I_i, y_j\right]^T, \tag{6}$$

where $I_i = iQ_I, i \in \{0, 1, ..., N_I\}$ and $y_j = jQ_y, j \in \{0, 1, ..., N_y\}$, and Q_I and Q_y are intervals for the intensity and the row directions, respectively, and $(N_I + 1)$ and $(N_y + 1)$ are the number of control points for the intensity and row directions, respectively (see Figure 3).

We then denote a response value $f_{i,j}$ be the transformed gait energy at the corresponding control point $\boldsymbol{r}_{i,j}$ of the i-th intensity value I_i and the j-th row y_j, where both the original value I_i and response value $f_{i,j}$ mean intensities (gait energies). In other words, given an intensity value I_i on the j-th row, it is transformed into the response value $f_{i,j}$. Similarly to the Global GERF, we represent the SD-GERF by concatenating the response values from all the control points to form a column vector $\boldsymbol{f} = \left[f_{0,0}, ..., f_{N_I,0}, \ldots, f_{0,N_y}, ..., f_{N_I,N_y}\right]^T \in \mathbb{R}^{N_d}$, $N_d = (N_I + 1)(N_y + 1)$.

We subsequently introduce the bilinear interpolation of response values for intermediate intensities and rows from the adjacent control points. Given a pixel with intensity I at the position (x, y), we compute an intensity index $i = \lfloor I/Q_I \rfloor$ and row index $j = \lfloor y/Q_y \rfloor$ of an adjacent control point, where $\lfloor \cdot \rfloor$ is a floor function. We then compute a corresponding response value $f(I, y)$ by bilinear interpolation from four pairs of adjacent control points and the corresponding response values, $(\boldsymbol{r}_{i,j}, f_{i,j})$, $(\boldsymbol{r}_{i,j+1}, f_{i,j+1})$, $(\boldsymbol{r}_{i+1,j}, f_{i+1,j})$, $(\boldsymbol{r}_{i+1,j+1}, f_{i+1,j+1})$ as

$$\begin{aligned} f(I, y) &= w_{i,j} f_{i,j} + w_{i+1,j} f_{i+1,j} + w_{i,j+1} f_{i,j+1} \\ &\quad + w_{i+1,j+1} f_{i+1,j+1} \\ &= \boldsymbol{w}_{I,y}^T \boldsymbol{f}. \end{aligned} \tag{7}$$

Here, $w_{i,j} = (1-w_I)(1-w_y)$, $w_{i+1,j} = w_I(1-w_y)$, $w_{i,j+1} = (1-w_I)w_y$, $w_{i+1,j+1} = w_I w_y$ are weights of individual adjacent control points, where $w_I = (I - iQ_I)/Q_I$ and $w_y = (y - jQ_y)/Q_y$. The vector $\boldsymbol{w}_{I,y}$ is a coefficient vector for the bilinear interpolation whose $i + j(N_I + 1)$-th, $(i+1) + j(N_I + 1)$-th, $i + (j+1)(N_I + 1)$-th, and $(i+1) + (j+1)(N_I + 1)$-th components are $w_{i,j}$, $w_{i+1,j}$, $w_{i,j+1}$, and $w_{i+1,j+1}$ and the other components are zero-padded.

Next, we measure a dissimilarity between a pair of transformed GEIs from original GEIs I_1 and I_2 via the SD-GERF $\boldsymbol{f}$ by a squared Euclidean distance as

$$\begin{aligned} d^2_{I_1,I_2} &= \sum_{x,y} \left(f(I_1(x,y),y) - f(I_2(x,y),y)\right)^2 \\ &= \sum_{x,y} \left(\boldsymbol{w}^T_{I_1(x,y),y}\boldsymbol{f} - \boldsymbol{w}^T_{I_2(x,y),y}\boldsymbol{f}\right)^2 \\ &= \boldsymbol{f}^T A_{I_1,I_2}\boldsymbol{f}, \end{aligned} \tag{8}$$

and the coefficient matrix $A_{I_1,I_2,} \in \mathbb{R}^{N_d \times N_d}$ can be computed as $A_{I_1,I_2} = \sum_{x,y} CC^T$, where $C = \boldsymbol{w}_{I_1(x,y),y} - \boldsymbol{w}_{I_2(x,y),y}$.

3.3. Training of GERF

We optimize Global GERF and SD-GERF on a training set that includes covariate variations to make the transformed GEI discriminative under covariate variations. Suppose the whole training set contains of two subsets $\mathcal{S}$ and $\mathcal{D}$, where the subset $\mathcal{S}$ and $\mathcal{D}$ are the sets of GEI pairs of the same and different subjects, respectively. To achieve better discrimination capability, it is preferable to decrease the sum of squared distances $D_\mathcal{S}$ for the same-subject pairs $\mathcal{S}$ while increasing the squared distances $D_\mathcal{D}$ for the different-subject pairs $\mathcal{D}$. Here, $D_\mathcal{S}$ and $D_\mathcal{D}$ are calculated as

$$\begin{aligned} D_\mathcal{S} &= \sum_{(I_1,I_2)\in\mathcal{S}} d^2_{I_1,I_2} &&= \boldsymbol{f}^T S_\mathcal{S}\boldsymbol{f} \\ D_\mathcal{D} &= \sum_{(I_1,I_2)\in\mathcal{D}} d^2_{I_1,I_2} &&= \boldsymbol{f}^T S_\mathcal{D}\boldsymbol{f}\ , \end{aligned} \tag{9}$$

where $S_\mathcal{S}$ and $S_\mathcal{D}$ are computed as $S_\mathcal{S} = \sum_{(I_1,I_2)\in\mathcal{S}} A_{I_1,I_2}$ and $S_\mathcal{D} = \sum_{(I_1,I_2)\in\mathcal{D}} A_{I_1,I_2}$, respectively.

In addition, a regularization term D_R is introduced to make the GERF smoother using first-order and second-order total variations [39], that is defined as

$$\begin{aligned} D_R &= w^I_1 \sum_{i=1}^{I_{\max}} (f_i - f_{i-1})^2 + w^I_2 \sum_{i=1}^{I_{\max}-1} (f_{i+1} - 2f_i + f_{i-1})^2 \\ &= \boldsymbol{f}^T(w^I_1 S_{R_1} + w^I_2 S_{R_2})\boldsymbol{f} \\ &= \boldsymbol{f}^T S_R \boldsymbol{f} \end{aligned} \tag{10}$$

for Global GERF, and defined as

$$
\begin{aligned}
D_R &= w_1^I \sum_{i=1}^{N_I} \sum_{j=0}^{N_y} (f_{i,j} - f_{i-1,j})^2 \\
&+ w_2^I \sum_{i=1}^{N_I-1} \sum_{j=0}^{N_y} (f_{i+1,j} - 2f_{i,j} + f_{i-1,j})^2 \\
&+ w_1^y \sum_{i=0}^{N_I} \sum_{j=1}^{N_y} (f_{i,j} - f_{i,j-1})^2 \\
&+ w_2^y \sum_{i=0}^{N_I} \sum_{j=1}^{N_y-1} (f_{i,j+1} - 2f_{i,j} + f_{i,j-1})^2 \\
&= \boldsymbol{f}^T (w_1^I S_{R_1}^I + w_2^I S_{R_2}^I + w_1^y S_{R_1}^y + w_4^y S_{R_2}^y) \boldsymbol{f} \\
&= \boldsymbol{f}^T S_R \boldsymbol{f}
\end{aligned}
\tag{11}
$$

for SD-GERF, where w_1^I, w_2^I, w_1^y and w_2^y are weighting parameters for the first-order and second-order smoothness terms for adjacent intensities and rows, respectively. The coefficient matrices $S_{R_1}^I$, $S_{R_2}^I$, $S_{R_1}^y$, and $S_{R_2}^y$ can be easily derived referring to the previous paper [20].

Finally, we optimize the GERF to maximize the Fisher ratio between the sum of squared distances $D_{\mathcal{D}}$ for the different-subject pairs and those $D_{\mathcal{S}}$ for the same-subject pairs plus the regularization term D_R under an L_2 norm constraint on $\boldsymbol{f}$ as

$$
\boldsymbol{f}^* = \arg\max_{\boldsymbol{f}} \frac{\boldsymbol{f}^T S_{\mathcal{D}} \boldsymbol{f}}{\boldsymbol{f}^T (S_{\mathcal{S}} + S_R) \boldsymbol{f}} \quad \text{s.t.} \quad \|\boldsymbol{f}\| = 1. \tag{12}
$$

We then formulate this optimization problem as the following generalized eigenvalue problem which is similar to the well-known LDA formulation,

$$
S_{\mathcal{D}} \boldsymbol{f} = \lambda (S_{\mathcal{S}} + S_R) \boldsymbol{f} \quad \text{s.t.} \quad \|\boldsymbol{f}\| = 1, \tag{13}
$$

where λ and $\boldsymbol{f}$ are an eigenvalue and its corresponding eigenvector, respectively. We therefore analytically obtain the optimal GERF $\boldsymbol{f}^*$ in a closed-form solution by assigning the eigenvector corresponding to the largest eigenvalue without any iterations.

3.4. Score-Level Fusion of Multiple SD-GERFs

We considered only the largest eigenvector as a solution of the GERF in our previous paper [20]. However, we can exploit multiple largest eigenvectors to obtain multiple GERFs for better accuracy in practice. Particularly, in the case of the SD-GERF, responses for each GERF often localize to a certain part. For example, the single largest eigenvector is localized to the head, and the second-largest is localized to the leg (refer to the feature examples in Section 4.4). Therefore, a single SD-GERF cannot reflect information from the whole body and may result in low recognition accuracy. We therefore adopt multiple SD-GERFs corresponding to the N largest eigenvalues to cover the whole body and fuse results from the multiple SD-GERFs for better accuracy.

As for the fusion scheme, we first compute dissimilarity scores for each SD-GERF between a probe and every subject in a gallery set and apply probe-dependent z-normalization [40] to the set of dissimilarity scores, i.e., linearly normalize the scores so that their mean and variance are 0 and 1, respectively. Once we obtain N z-normalized dissimilarity scores from N SD-GERFs, we fuse them by either of two simple, yet effective, score-level fusion methods [41], i.e., sum and min rule.

3.5. Post-Processing

After transforming the GEI with the optimal GERF, we adopt two sequential processes for further improvement. The first one is Gabor filtering, which has been demonstrated to be effective for

gait recognition and Gabor functions-based image decomposition is biologically relevant to image understanding and recognition as reported in [18,42]. Similar to [42], we use subsampled Gabor features with Gabor functions from five scales and eight orientations. Because the silhouette gait feature provided in the databases is 88×128, the resolution of the Gabor feature is 320×352. An example of Global GERF after Gabor filtering is shown in Figure 4. The second one is a spatial metric learning, that is, two-dimensional PCA (2DPCA) [43] and two-dimensional LDA (2DLDA). They are used not only for reducing feature dimensions, but also for getting more discriminant features. Similar to [26,27], they are applied in Gabor feature space for horizontal and vertical direction, respectively.

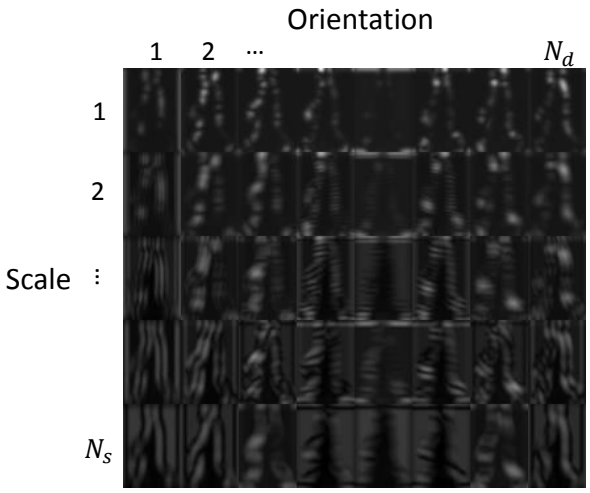

Figure 4. An example of Global GERF after Gabor filtering. The rows show different scales and the columns show different orientations. In this figure, the number of scales is $N_s = 5$ and the number of orientations is $N_d = 8$, respectively.

Next, the dissimilarity score between the gallery and probe elements of a pair is measured as the Euclidean distance in the 2DPCA+2DLDA space. Finally, in verification scenarios, the score is compared with an acceptance threshold to verify whether the pair belongs to the same subject or not. While in identification scenarios, nearest neighbor classifier is used to assign the final identity of the probe according to the dissimilarity scores between the probe and all the galleries.

4. Experiments

4.1. Datasets

We used three databases (OU-TD-B and OU-LP-Bag β databases are available at http://www.am.sanken.osaka-u.ac.jp/BiometricDB/index.html, CASIA-B database is available at http://www.cbsr.ia.ac.cn/english/Gait%20Databases.asp), i.e., the OU-ISIR Gait Database, the Treadmill Dataset B [21] (OU-TD-B), the OU-ISIR Gait database, Large Population dataset with bag, β version [22] (OU-LP-Bag β) and the CASIA Gait Database B (CASIA-B) [23] for the experiments.

The OU-TD-B has the largest number of clothing variations, at most 32. It is separated into three subsets. The training set contains 446 sequences of 20 subjects with the range of 15 to 28 different combinations of clothing. The gallery and probe sets constitute a testing set that comprises 48 subjects, which are disjoint from the training set. Standard clothing type (i.e., regular pants and full shirt) is only included in the gallery set, while other clothing types with 856 sequences are included in the probe set.

The OU-LP-Bag β includes various carrying statuses in the wild. There are 2070 subjects in the dataset and each subject has two sequences, one with carried objects and the other without carried objects. The whole dataset has three subsets: a training set, a gallery set, and a probe set. There are 2068 sequences of 1034 subjects in the training set, while the remaining 1036 subjects are included in

the gallery and probe sets. The sequences in the gallery set have no carried objects, while the sequences in the probe set have carried objects.

The CASIA-B contains 124 subjects from 11 views (from 0 to 180 degree with an interval of 18 degree). For each view, there are ten sequences per subject. Six of them are captured under normal walking conditions (NM); two of them are captured when carrying a bag (BG); the rest two are captured when wearing a coat (CL). Among the ten sequences, the first four sequences under normal walking conditions are chosen as the gallery (NM #1-4). The other six sequences are kept as three probe sets under different walking conditions: (1) Set-A contains two NM sequences (NM #5-6); (2) Set-B contains two BG sequences (BG #1-2); (3) Set-C contains two CL sequences (CL #1-2).

4.2. Parameter Setting

The proposed method has several hyper-parameters, i.e., weighting coefficients of the regularization terms both in intensity and spatial direction (w_1^I, w_2^I, w_1^y and w_2^y), and the 2DLDA dimension d (corresponding to the chosen number of largest eigenvalues in 2DLDA). We experimentally set $w_1^I = w_2^I$, $w_1^y = w_2^y$, and changed them in the range of $[10, 10^2, \ldots, 10^5]$. We also changed 2DLDA dimension d in the range of $[10, 20, \ldots, 200]$. All the hyper-parameters were automatically selected by a grid search for each verification and identification scenarios; i.e., we adopted the hyper-parameters that achieved the best accuracy on the training set regarding verification and identification criteria, respectively. In addition, we had two choices for the score-level fusion methods, i.e., sum or min rule, and we adopted the better one, similarly to the previously mentioned hyper-parameter selection process. Moreover, the SD-GERF still has some parameters, namely, intervals for intensity Q_I, that for row Q_y, and the number of multiple eigenvectors (i.e., response functions) N. For these parameters, we experimentally set $Q_I = 16$, $Q_y = 8$, and $N = 10$ and used them throughout all the experiments. Regarding the parameters in post-processing, we set Gabor window size to 23×23 (Note that we used a different Gabor window size from the previous work [20] for the sake of balancing accuracy and computational time.) and keep 99% variance for 2DPCA.

4.3. Evaluation Metrics

We evaluated the recognition accuracy of the proposed method both in verification and identification scenarios. In verification scenarios, a detection error tradeoff (DET) curve is employed that indicates a tradeoff between false non-match rate (FNMR) and false match rate (FMR) when an acceptance threshold changes. Specifically, FNMR is the proportion of genuine attempts that are falsely declared not to match a template of the same subject and FMR is the proportion of the imposter attempts that are falsely declared to match a template of another subject. In addition, an equal error rate (EER), where FNMR is equal to FMR, is also presented. In identification scenarios, a cumulative match characteristic (CMC) curve is employed that shows identification rates of actual subjects included within each of the ranks. In addition, the rank-1 identification rate is also shown. Obviously, the lower EER value and higher rank-1 identification rate mean the higher recognition performance.

4.4. Comparison with Intensity Transformation-Based Methods

We conducted comparison experiments with a family of intensity transformation-based methods including GEnI [12], Masked GEI [13] and GEI [11] as a baseline to show the effectiveness of the proposed data-driven transformation methods Global GERF and SD-GERF.

We first show some feature examples for the five gait features, that is, GEI, GEnI, Masked GEI, Global GERF and SD-GERF, along with profiles of three SD-GERFs corresponding to the three largest eigenvalues in Figure 5. Note that the proposed SD-GERF uses 10 eigenvectors as multi-response functions; here, we only list three of them as examples. As for the profile of Global GERF, it is drawn as a red curve in the third column of Figure 1.

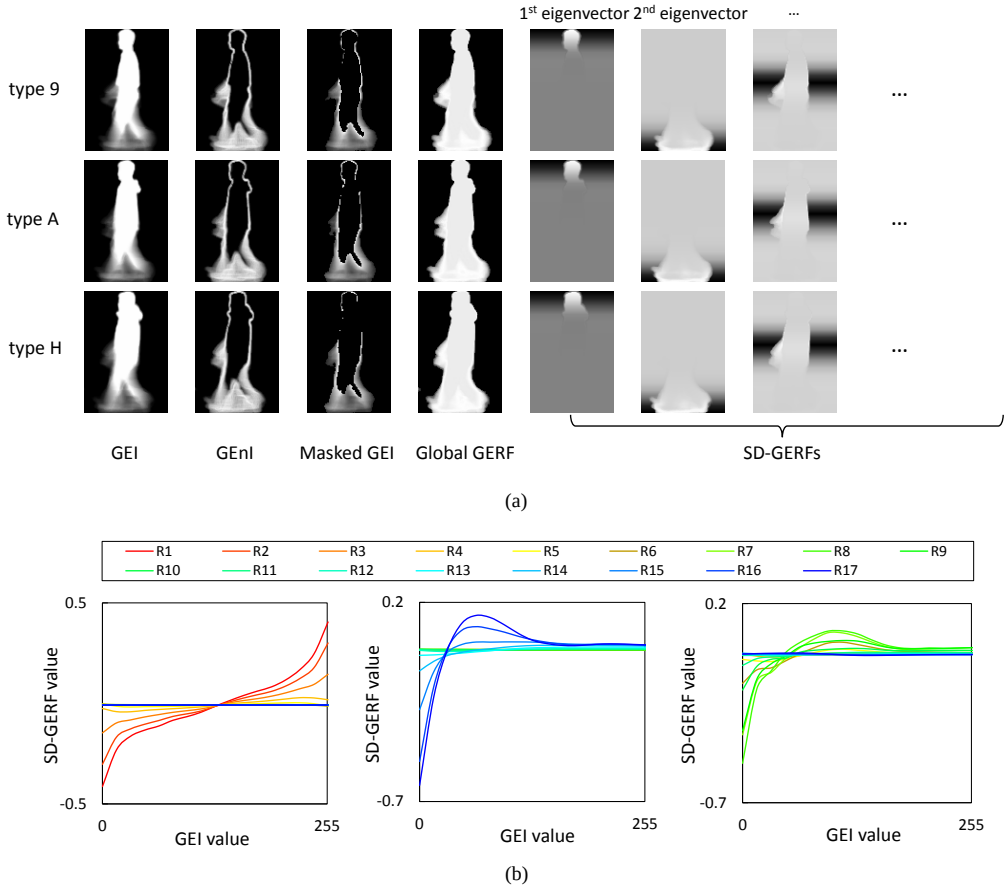

Figure 5. Examples of features and profiles of the SD-GERF in OU-TD-B. (**a**) Features of five intensity transformation-based methods. Three features corresponding to the three largest eigenvalues are listed for the SD-GERF as examples. (**b**) Profiles of response functions of the SD-GERF corresponding to the three largest eigenvalues. R1, R2, . . . , R17 indicate row indices from the top to the bottom, where the control points are aligned. Note that some profiles of the SD-GERFs are not shown because they are overlapped with those of similar SD-GERFs.

The profile reveals that on one hand the Global GERF compresses the intensity difference between typical dynamic components (intensity values around 127) and the complete foreground (intensity values around 255) to mitigate the effects of clothing variations, e.g., the leg is usually occupied by the typical dynamic components in the case of standard clothing, while sometimes occupied by the complete foreground in the case of long coats or skirts. On the other hand, it simultaneously emphasizes the difference between the complete background (intensity values around 0) and the typical dynamic components, because such differences are still useful under clothing variations. In addition, it also keeps the difference between complete background and foreground as this still contains meaningful information to discriminate subjects under conditions of clothing variations, unlike GEnI or Masked GEI, which make no such distinction.

As for the SD-GERF, it exhibits similar trends in profile to those of the Global GERF, as shown in the second column (localized to the leg) and third column (localized to the torso) of Figure 5b. This implies that the SD-GERF highlights differences in dynamic components and retains static components such as the Global GERF. However, the first column (localized to the head) in Figure 5b shows monotonically increasing profiles unlike those for the leg and the torso (the second and third columns). This is because the head is not often affected by clothing variations and hence the original GEI is sufficient for discrimination. As such, the SD-GERF can consider the spatial dependency on the degree of effect of clothing variations and thus set appropriate response functions for each body part.

The DET curves with z-normalization and CMC curves of all the transformation-based methods are shown in Figure 6. In addition, the detailed EER with z-normalization and rank-1 identification rate are listed in Table 1. From the results, the proposed GERFs (Global GERF and SD-GERF) outperform the other three transformation-based methods, in the case of both verification and identification. Moreover, the SD-GERF achieves better accuracy than the Global GERF. For instance, the 4.27% lower z-EER and 10.9% higher rank-1 rate in OU-TD-B indicate that addition of spatial dependency in the GERF framework improves accuracy.

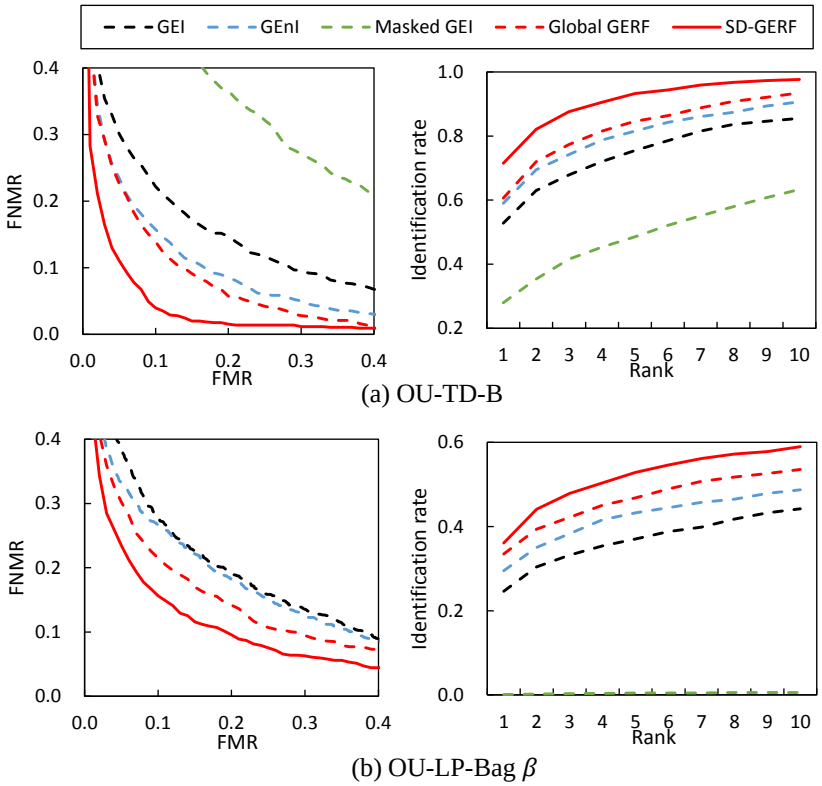

Figure 6. Detection error tradeoff (DET) curves with z-normalization (**left**) and cumulative match characteristic (CMC) curves (**right**) for intensity transformation-based methods in two datasets.

Table 1. Equal error rate (EER) with z-normalization (denoted as z-EER) [%] and rank-1 identification rate (denoted as Rank-1) [%] for intensity transformation-based methods. Digits with bold and italic bold fonts represent the best and the second-best results, separately. This convention is consistent throughout the whole paper.

Dataset	**OU-TD-B**		**OU-LP-Bag β**	
Methods	**z-EER**	**Rank-1**	**z-EER**	**Rank-1**
GEI [11]	16.12	52.8	19.59	24.6
GEnI [12]	12.81	59.0	18.82	29.5
Masked GEI [13]	28.15	28.0	61.95	0.1
Global GERF [20]	***11.68***	***60.6***	***16.22***	***33.5***
SD-GERF (proposed)	**7.41**	**71.5**	**12.99**	**36.1**

4.5. Comparison on Clothing and Carrying Status Variations

In this subsection, we evaluated the proposed methods (combined with two post-processing techniques, denoted as Gabor+Global GERF [20] and Gabor+SD-GERF) against clothing and carrying status variations on OU-TD-B, OU-LP-Bag β and CASIA-B (in case of side-view). OU-TD-B and

OU-LP-Bag β are more difficult compared with CASIA-B since they have more variations, thus they are detailed analyzed both in identification and verification scenarios.

4.5.1. OU-TD-B and OU-LP-Bag β

The state-of-the-art methods for comparison contain SVB-frieze pattern [24], Component-based [44], Whole-based [16], Part-based [28], Part-EnDFT [29], AESI + ZNK [45], GEI+RSM [26], Gabor+RSM-HDF [27], Gabor GEI [18], two-point gait (TPG) + GEI [46], GEI w/LDA [47], GEI w/2DLDA [48], GEI w/CSA [49], GEI w/DATER [25], GEI w/Ranking SVM [50] and JIS-ML [22]. Moreover, we evaluated a CNN-based method, i.e., GEINet [34] as one of the benchmarks to make a comparison with deep learning-based methods. We trained the GEINet using the same dataset protocol as our GERF model. All default hyper-parameters are chosen except for the number of units on the fully connected layer fc4, which is equal to the number of subjects in the training set. Thus, we changed it according to our datasets. Finally, the DET curves with z-normalization and CMC curves of all the methods are shown in Figure 7 and EERs with z-normalization and rank-1 identification rates are shown in Table 2.

Table 2. EER with z-normalization [%] and rank-1 identification rates [%] compared with the state-of-the-art methods. N/A and "-" mean not applicable and not provided, respectively.

Dataset	**OU-TD-B**		**OU-LP-Bag β**	
Methods	**z-EER**	**Rank-1**	**z-EER**	**Rank-1**
SVB-frieze pattern [24]	19.81	-	-	-
Component-based [44]	18.25	-	-	-
Whole-based [16]	14.88	58.1	-	-
Part-based [28]	10.26	66.3	-	-
Part-EnDFT [29]	-	72.8	-	-
AESI+ZNK [45]	-	72.7	-	-
GEI+RSM [26]	N/A	80.4	N/A	N/A
Gabor+RSM-HDF [27]	N/A	**90.7**	N/A	N/A
Gabor GEI [18]	11.80	62.3	10.48	46.4
TPG+GEI [46]	7.10	-	-	-
GEI w/LDA [47]	15.63	54.3	8.10	54.6
GEI w/2DLDA [48]	8.91	70.7	11.47	43.3
GEI w/CSA [49]	16.00	-	-	-
GEI w/DATER [25]	8.72	-	-	-
GEI w/Ranking SVM [50]	10.75	58.4	10.81	28.3
JIS-ML [22]	6.66	74.5	**5.45**	57.4
GEINet [34]	8.38	60.2	9.75	40.7
Gabor+Global GERF [20]	***5.14***	82.7	6.67	***58.3***
Gabor+SD-GERF (proposed)	**4.61**	***87.4***	***5.60***	**64.3**

The proposed Gabor+SD-GERF method achieves the best or second-best performance for both datasets, which shows that this method is effective and robust for gait recognition under variations in clothing and carrying status. Although Gabor+RSM-HDF [27] got the best rank-1 identification rate in OU-TD-B, we must point out that the RSM framework suffers from three weaknesses: (1) It can be only applied to identification scenarios (not to verification scenarios) because it relies on majority voting in all the galleries; (2) Because it requires multiple samples per gallery to compute a within-class scatter from the gallery set, it cannot be applied to datasets with a single sample per gallery (e.g., OU-LP-Bag β); and (3) It cannot guarantee reproducibility because it contains a random selection process at the metric learning stage. Moreover, compared with the joint intensity transformation-based method (JIS-ML [22]), the proposed method achieves higher performance in both datasets except for a slightly lower z-EER (0.15%) in OU-LP Bag β.

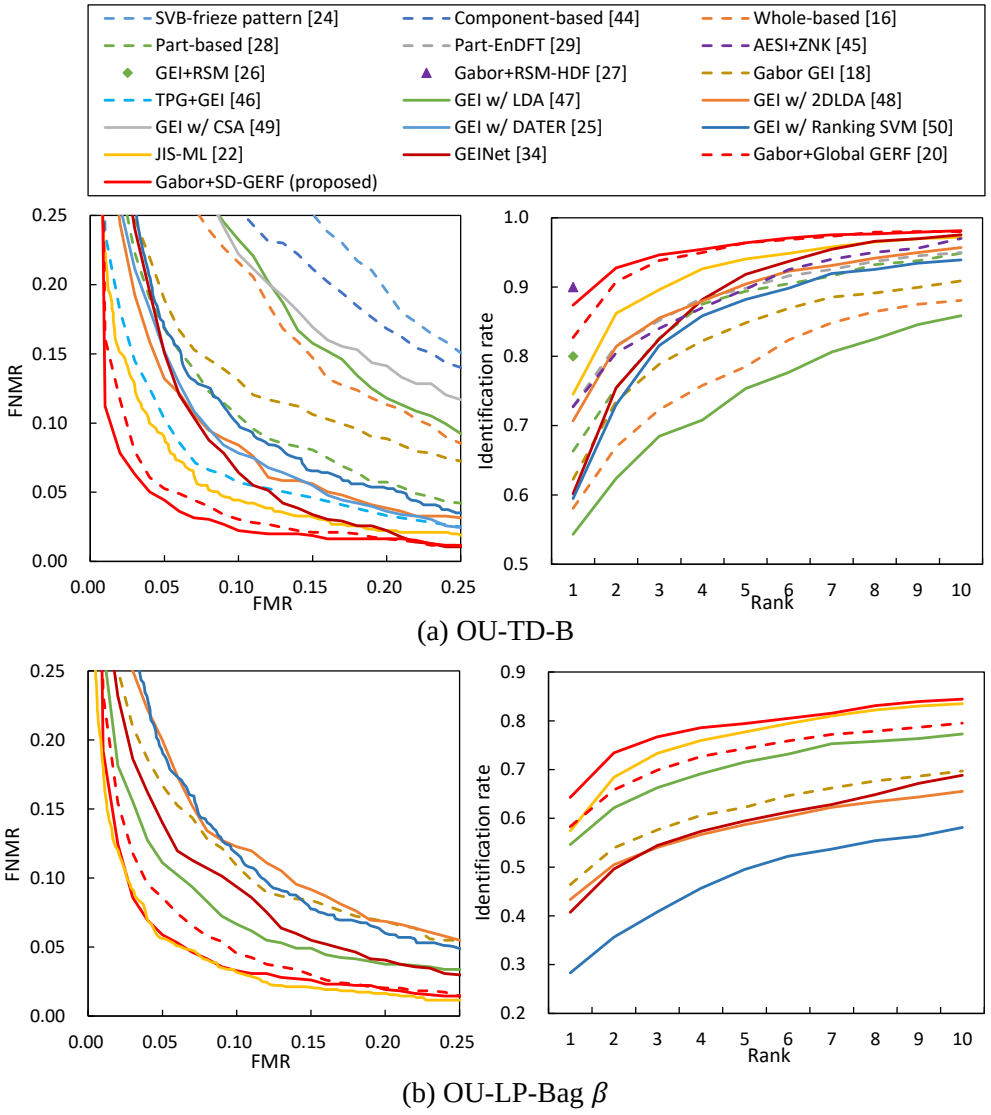

Figure 7. DET curves with z-normalization (**left**) and CMC curves (**right**) for comparison with state-of-the-art methods in two datasets.

Even compared with GEINet, the proposed method achieves better performance. This is because that the GEINet was firstly designed for cross-view gait recognition on the OU-ISIR large population dataset (OULP) [51] which uses more than 15,000 samples. When compared with this large population dataset, the clothing and carrying status datasets we used in our experiments (about 1300 samples for OU-TD-B and 4000 samples for OU-LP-Bag β) are quite small although they have the largest clothing and carrying status variations among publicly available gait datasets. Such small datasets cause the overfitting problem and hence the GEINet suffers from large generalization errors. On the other hand, the number of clothing and carrying status variations (e.g., 32 clothing types and unpredictable carrying status including various location and size) is much larger than the number of view variations (only 4 view angles in OULP), which increases the difficulty of recognition. Therefore, the GEINet is inferior to the proposed method unless using a much larger training set, or even improving the network architecture.

Considering the aforementioned aspects, we argue that the proposed method is superior to the other benchmarks because of its wider application range and very competitive accuracies. Upon closer examination of an extended part of the GERF framework, Gabor+SD-GERF gets better results than Gabor+Global GERF (The z-EER and rank-1 are slightly different from those in the previous paper [20]

because different hyper-parameters were automatically chosen based on the training set in this paper.) and better than the case without post-processing in Section 4.4, from Table 2 0.53% lower z-EER and 4.7% higher rank-1 in OU-TD-B, and 1.07% lower z-EER and 6.0% higher rank-1 in OU-LP-Bag β. All these results show the effectiveness of the proposed spatially dependent framework of the GERF.

4.5.2. CASIA-B

We used the side-view sequences in CASIA-B to keep the view angle unchanged for evaluation under different walking conditions. For each probe set, we trained one GERF model using the first 24 subjects, while the rest 100 subjects are used for testing. Only the performance in identification scenarios is shown for comparison since few works reported their results in verification scenarios. Because we always got 100% rank-1 rate on the training set due to the small number of training samples and variations of each walking condition, it is hard to choose the hyper-parameters based on the training set. Thus, the hyper-parameters are set to be the same as those in OU-TD-B. Table 3 shows the rank-1 identification rates. The compared results were drawn from original papers, except for GEINet [34] which was reported in [52]. From the results, the proposed method significantly outperforms the other existing methods, especially for Set-B and Set-C which contain BG and CL, which show the effectiveness of the proposed method tackling clothing and carrying status variations. Two CNN-based methods, i.e., GEINet [34] and DCNN [53], fail to achieve higher performance due to the small number of training samples. Also, when compared with the extended part of the GERF framework, Gabor+SD-GERF gets better results than Gabor+Global GERF which shows the effectiveness of the proposed spatially dependent framework of the GERF.

Table 3. Side-view rank-1 identification rates [%] compared with other state-of-the-arts on three probe sets under different walking conditions of CASIA-B dataset.

Methods	Set-A	Set-B	Set-C	Average
GEI [11]	***99***	60	30	63.0
GEnI [12]	98.3	80.1	33.5	70.6
STIP+NN [54]	95.4	60.9	52	69.4
AESI+ZNK [45]	**100**	***93.1***	81.3	91.5
L-CRF [52]	98.6	90.2	85.8	91.5
GEINet [34]	97.5	84.5	71.8	84.6
DCNN [53]	95.6	88.3	76.2	86.7
Gabor+Global GERF [20]	***99***	91	***92***	***94.0***
Gabor+SD-GERF (proposed)	***99***	**100**	**96**	**98.3**

4.6. Discussion

4.6.1. Comparison on Speed Variation

We discussed the performance of proposed method on speed variation, which involves different covariate conditions compared with clothing and carrying status. While the clothing and carrying status variations tend to affect the static components (e.g., torso and limb shapes) more than the dynamic ones (e.g., changes in stride and arm swing), speed variation displays the opposite pattern. The OU-ISIR Gait Database, Treadmill Dataset A [21] (OU-TD-A) is used for this experiment, because it has the largest speed variation among publicly available gait databases. We trained a common GERF model for all speed cases. Table 4 shows the rank-1 identification rates. From the results, the proposed method achieves the second-best performance. However, it is noted that SSGEI [55] which achieves the best performance is specifically designed for solving the speed variation.

We also made some qualitative evaluation in Figure 8, which shows typical feature examples and profile of Global GERF in OU-TD-A. We find that the obtained Global GERF (bottom row of Figure 8a) is more insensitive to the difference between dynamic components of the legs and arms and

the background, than the original GEI (top row of Figure 8a). This is also understandable from the profile shown in Figure 8b. Intensity differences between background (intensity values around 0) and typical dynamic components (intensity values around 127) are compressed (i.e., nearly flat), and hence the dynamic components approach background intensity. In contrast, intensity differences between the typical dynamic components and foreground (intensity values around 255) are emphasized. This profile suggests a benefit of discriminating foreground from the others, regardless of whether they are background or dynamic components. Hence, the proposed GERF can attenuate the dynamic components, which are greatly affected by speed variation, and simultaneously highlight the differences in static components.

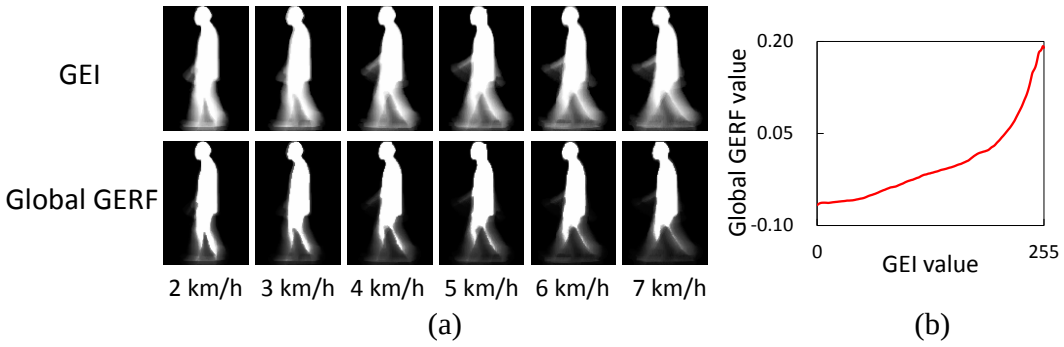

Figure 8. The Global GERF under speed variation: (**a**) feature examples of GEI and Global GERF under different speed (from 2 km/h to 7 km/h); (**b**) profile of the Global GERF.

Table 4. Rank-1 identification rates [%] compared with other state-of-the-arts in case of small speed change (3 vs. 4 km/h), large speed change (2 vs. 6 km/h) and average of all speed changes in OU-TD-A.

Methods	Small Speed	Large Speed	Average (All Speed)
STM [56]	90	58	-
DCM [57]	***98***	82	92.44
RSM [58]	**100**	95	98.07
SSGEI [55]	**100**	**98**	**99.33**
Gabor+Global GERF [20]	**100**	92	96.89
Gabor+SD-GERF (proposed)	**100**	***96***	***98.11***

Through this discussion of speed variation, we confirm the flexibility of the proposed GERF framework, in which the response functions are learned in a data-driven way, and show the potential for wide application of robust gait recognition under various covariate conditions, including different clothing, carrying status, and speed.

4.6.2. Consideration for Real-World Applications

Although we have achieved promising results under various covariate conditions, there still exists some limitations of the proposed method to real-world applications (e.g., surveillance), that is, the proposed method is based on the transformation of appearance-based features (i.e., GEI), whose quality is largely subject to the extracted silhouettes of a gait period. In case of real scenes, however, more difficult variations, such as complex background and occlusions, will greatly affect the extraction of subjects' silhouettes and may further drop the performance of gait recognition. Fortunately, recent deep learning-based techniques (e.g., Mask R-CNN [59]) have brought great improvement in the performance of human detection and segmentation in complex real scenes, which benefits for the silhouette extraction to generate good appearance-based features. This implies a combination of these methods for real scene applications.

4.7. Evaluation of Computational Time

We ran the MATLAB code of the proposed method on a PC with an Intel Core i7 4.00 GHz processor and 32 GB RAM to evaluate the computational time. For OU-TD-B dataset, we show the training time of SD-GERF and spatial metric learning, together with the query time of each sequence based on GEI templates in Table 5. One of the benchmarks, i.e., Gabor+RSM-HDF [27], is used for comparison, which analyzed the computational time on USF dataset [9] including 122 subjects in the gallery set. For fair comparison, we further estimate the proposed method under a comparable setting by considering the computing power of the computer and the number of gallery sequences. Note that the running time excludes the computation of Gabor features similarly to [27], which costs about 0.05 s per GEI. As a result, the proposed method shows lower computational time than Gabor+RSM-HDF and thus is more suitable for real applications.

Table 5. Running time (Seconds) comparison.

Method \ Running Stage	Machine Specification	Training Time	Query Time of Each Sequence (#Gallery Sequences)
Gabor+RSM-HDF [27]	Intel Core i5 3.10 GHz processor	320.090	0.600 (122)
Proposed method	Intel Core i7 4.00 GHz processor	13.330	0.016 (48)
Proposed method (estimated)	75% computing power	17.773	0.054 (122)

5. Conclusions

In this paper, we described a data-driven framework to learn GERFs (i.e., Global GEI and SD-GERF) for gait recognition against clothing and carrying status. Specifically, we first proposed the Global GERF to transform an original gait energy into another value to make it more discriminative under variations. In addition, since the discrimination capability of gait energies, as well as the degree to which they are affected by the covariate conditions, differs among body parts, we then extended the Global GERF to SD-GERF which accounts for spatial dependence. Moreover, the proposed GERFs were represented as a lookup table vector and optimized through an efficient generalized eigenvalue problem, which enables us to obtain an analytical solution in a closed form without any iterations. To further improve accuracy, two post-processing techniques (i.e., Gabor filtering and 2DPCA+2DLDA) were employed. Experimental results using three publicly available datasets showed the state-of-the-art performance of the proposed GERF compared with other state-of-the-art methods.

Because the proposed framework can be regarded as a kind of feature learning method, we will apply the method to other problems such as gait-based gender and age estimation or even more dissimilar fields where intensity plays an important role in comparison, such as face recognition. Moreover, considering the good performance of deep learning-based methods, integrating the GERF model into a deep network architecture will remain as another future work.

Author Contributions: X.L. developed the program for the proposed algorithm, performed most of the experiments and drafted the initial manuscript. Y.M. guided the paper's content, including the idea, experimental design, data analysis, as well as initial manuscript revision. C.X. participated in the implementation of post-processing techniques and provided part of the experimental results. Y.M. and D.M. developed the main idea and equations of the proposed method. Y.Y. and M.R. supervised the work providing technique support and general guidance of the whole work. All authors participated in review of the final manuscript.

Funding: This work was supported by JSPS Grants-in-Aid for Scientific Research (A) 18H04115, the National R&D Program for Major Research Instruments (Grant No. 61727802), the National Natural Science Foundation of China (Grants No. 61703209).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Nixon, M.S.; Tan, T.; Chellappa, R. *Human Identification Based on Gait*; International Series on Biometrics; Springer: Berlin/Heidelberg, Germany, 2005.
2. Bouchrika, I.; Goffredo, M.; Carter, J.; Nixon, M. On Using Gait in Forensic Biometrics. *J. Forensic Sci.* **2011**, *56*, 882–889. [CrossRef] [PubMed]
3. Iwama, H.; Muramatsu, D.; Makihara, Y.; Yagi, Y. Gait Verification System for Criminal Investigation. *IPSJ Trans. Comput. Vis. Appl.* **2013**, *5*, 163–175. [CrossRef]
4. Lynnerup, N.; Larsen, P. Gait as evidence. *IET Biom.* **2014**, *3*, 47–54. [CrossRef]
5. Urtasun, R.; Fua, P. 3D Tracking for Gait Characterization and Recognition. In Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition, Seoul, Korea, 17–19 May 2004; pp. 17–22.
6. Wagg, D.; Nixon, M. On Automated Model-Based Extraction and Analysis of Gait. In Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition, Seoul, Korea, 17–19 May 2004; pp. 11–16.
7. Yam, C.; Nixon, M.; Carter, J. Automated Person Recognition by Walking and Running via Model-based Approaches. *Pattern Recognit.* **2004**, *37*, 1057–1072. [CrossRef]
8. Zhao, G.; Liu, G.; Li, H.; Pietikainen, M. 3D gait recognition using multiple cameras. In Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition, Southampton, UK, 10–12 April 2006; pp. 529–534.
9. Sarkar, S.; Phillips, J.; Liu, Z.; Vega, I.; Grother, P.; Bowyer, K. The HumanID Gait Challenge Problem: Data Sets, Performance, and Analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 162–177. [CrossRef] [PubMed]
10. Wang, L.; Tan, T.; Ning, H.; Hu, W. Silhouette analysis-based gait recognition for human identification. *IEEE Trans. Pattern Anal. Mach. Intell.* **2003**, *25*, 1505–1518. [CrossRef]
11. Han, J.; Bhanu, B. Individual Recognition Using Gait Energy Image. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 316–322. [CrossRef] [PubMed]
12. Bashir, K.; Xiang, T.; Gong, S. Gait recognition using gait entropy image. In Proceedings of the 3rd International Conference on Imaging for Crime Detection and Prevention (ICDP), London, UK, 3 December 2009; pp. 1–6.
13. Bashir, K.; Xiang, T.; Gong, S. Gait recognition without subject cooperation. *Pattern Recognit. Lett.* **2010**, *31*, 2052–2060. [CrossRef]
14. Castro, F.M.; Marín-Jiménez, M.J.; Mata, N.G.; Muñoz-Salinas, R. Fisher Motion Descriptor for Multiview Gait Recognition. *Int. J. Pattern Recognit. Artif. Intell.* **2017**, *31*, 1–40. [CrossRef]
15. Marín-Jiménez, M.J.; Castro, F.M.; Mata, N.G.; de la Torre, F.; Muñoz-Salinas, R. Deep Multi-Task Learning for Gait-based Biometrics. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 106–110.
16. Makihara, Y.; Sagawa, R.; Mukaigawa, Y.; Echigo, T.; Yagi, Y. Gait Recognition Using a View Transformation Model in the Frequency Domain. In Proceedings of the 9th European Conference on Computer Vision (ECCV), Graz, Austria, 7–13 May 2006; pp. 151–163.
17. Wang, C.; Zhang, J.; Wang, L.; Pu, J.; Yuan, X. Human Identification Using Temporal Information Preserving Gait Template. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 2164–2176. [CrossRef] [PubMed]
18. Tao, D.; Li, X.; Wu, X.; Maybank, S.J. General Tensor Discriminant Analysis and Gabor Features for Gait Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 1700–1715. [CrossRef] [PubMed]
19. Liu, Z.; Sarkar, S. Simplest Representation Yet for Gait Recognition: Averaged Silhouette. In Proceedings of the 17th International Conference on Pattern Recognition (ICPR), Cambridge, UK, 26 August 2004; Volume 4, pp. 211–214.
20. Li, X.; Makihara, Y.; Xu, C.; Muramatsu, D.; Yagi, Y.; Ren, M. Gait Energy Response Function for Clothing-Invariant Gait Recognition. In Proceedings of the 13th Asian Conference on Computer Vision (ACCV), Taipei, Taiwan, 20–24 November 2016; pp. 257–272.
21. Makihara, Y.; Mannami, H.; Tsuji, A.; Hossain, M.; Sugiura, K.; Mori, A.; Yagi, Y. The OU-ISIR Gait Database Comprising the Treadmill Dataset. *IPSJ Trans. Comput. Vis. Appl.* **2012**, *4*, 53–62. [CrossRef]

22. Makihara, Y.; Suzuki, A.; Muramatsu, D.; Li, X.; Yagi, Y. Joint Intensity and Spatial Metric Learning for Robust Gait Recognition. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 5705–5715.
23. Yu, S.; Tan, D.; Tan, T. A Framework for Evaluating the Effect of View Angle, Clothing and Carrying Condition on Gait Recognition. In Proceedings of the 18th International Conference on Pattern Recognition (ICPR), Hong Kong, China, 20–24 August 2006; Volume 4, pp. 441–444.
24. Lee, S.; Liu, Y.; Collins, R. Shape Variation-Based Frieze Pattern for Robust Gait Recognition. In Proceedings of the International Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007; pp. 1–8.
25. Xu, D.; Yan, S.; Tao, D.; Zhang, L.; Li, X.; Zhang, H.J. Human gait recognition with matrix representation. *IEEE Trans. Circuits Syst. Video Technol.* **2006**, *16*, 896–903.
26. Guan, Y.; Li, C.T.; Hu, Y. Robust Clothing-Invariant Gait Recognition. In Proceedings of the Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), Piraeus-Athens, Greece, 18–20 July 2012; pp. 321–324.
27. Guan, Y.; Li, C.T.; Roli, F. On Reducing the Effect of Covariate Factors in Gait Recognition: A Classifier Ensemble Method. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1521–1528. [CrossRef] [PubMed]
28. Hossain, M.A.; Makihara, Y.; Wang, J.; Yagi, Y. Clothing-Invariant Gait Identification using Part-based Clothing Categorization and Adaptive Weight Control. *Pattern Recognit.* **2010**, *43*, 2281–2291. [CrossRef]
29. Rokanujjaman, M.; Islam, M.; Hossain, M.; Islam, M.; Makihara, Y.; Yagi, Y. Effective Part-Based Gait Identification using Frequency-Domain Gait Entrophy Features. *Multimedia Tools Appl.* **2015**, *74*, 3099–3120. [CrossRef]
30. Boulgouris, N.; Chi, Z. Human gait recognition based on matching of body components. *Pattern Recognit.* **2007**, *40*, 1763–1770. [CrossRef]
31. Iwashita, Y.; Uchino, K.; Kurazume, R. Gait-Based Person Identification Robust to Changes in Appearance. *Sensors* **2013**, *13*, 7884–7901. [CrossRef] [PubMed]
32. Wu, Z.; Huang, Y.; Wang, L. Learning Representative Deep Features for Image Set Analysis. *IEEE Trans. Multimedia* **2015**, *17*, 1960–1968. [CrossRef]
33. Wolf, T.; Babaee, M.; Rigoll, G. Multi-view gait recognition using 3D convolutional neural networks. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 4165–4169.
34. Shiraga, K.; Makihara, Y.; Muramatsu, D.; Echigo, T.; Yagi, Y. GEINet: View-Invariant Gait Recognition Using a Convolutional Neural Network. In Proceedings of the 8th IAPR International Conference on Biometrics (ICB), Halmstad, Sweden, 13–16 June 2016; pp. 1–8.
35. Zhang, C.; Liu, W.; Ma, H.; Fu, H. Siamese neural network based gait recognition for human identification. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 20–25 March 2016; pp. 2832–2836.
36. Wu, Z.; Huang, Y.; Wang, L.; Wang, X.; Tan, T. A Comprehensive Study on Cross-View Gait Based Human Identification with Deep CNNs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 209–226. [CrossRef] [PubMed]
37. Castro, F.M.; Marín-Jiménez, M.J.; Guil, N.; López-Tapia, S.; de la Blanca, N.P. Evaluation of CNN Architectures for Gait Recognition Based on Optical Flow Maps. In Proceedings of the International Conference of the Biometrics Special Interest Group (BIOSIG), Darmstadt, Germany, 20–22 September 2017; pp. 1–5.
38. Makihara, Y.; Yagi, Y. Silhouette Extraction Based on Iterative Spatio-Temporal Local Color Transformation and Graph-Cut Segmentation. In Proceedings of the 19th International Conference on Pattern Recognition (ICPR), Tampa, FL, USA, 8–11 December 2008; pp. 1–4.
39. Papafitsoros, K.; Schönlieb, C.B. A Combined First and Second Order Variational Approach for Image Reconstruction. *J. Math. Imaging Vis.* **2014**, *48*, 308–338. [CrossRef]
40. Phillips, P.; Blackburn, D.; Bone, M.; Grother, P.; Micheals, R.; Tabassi, E. Face Recogntion Vendor Test. 2002. Available online: https://www.nist.gov/itl/iad/image-group/face-recognition-vendor-test-frvt-2002 (accessed on 5 August 2017).
41. Kittler, J.; Hatef, M.; Duin, R.P.W.; Matas, J. On Combining Classifiers. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 226–239. [CrossRef]

42. Xu, D.; Huang, Y.; Zeng, Z.; Xu, X. Human Gait Recognition Using Patch Distribution Feature and Locality-Constrained Group Sparse Representation. *IEEE Trans. Image Process.* **2012**, *21*, 316–326. [CrossRef] [PubMed]
43. Yang, J.; Zhang, D.; Frangi, A.F.; Yang, J.Y. Two-dimensional PCA: A new approach to appearance-based face representation and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 131–137. [CrossRef] [PubMed]
44. Li, X.; Maybank, S.; Yan, S.; Tao, D.; Xu, D. Gait Components and Their Application to Gender Recognition. *IEEE Trans. Syst. Man Cybern. Part C* **2008**, *38*, 145–155.
45. Aggarwal, H.; Vishwakarma, D. Covariate conscious approach for Gait recognition based upon Zernike moment invariants. *IEEE Trans. Cogn. Dev. Syst.* **2018**, *10*, 397–407. [CrossRef]
46. Lombardi, S.; Nishino, K.; Makihara, Y.; Yagi, Y. Two-Point Gait: Decoupling Gait from Body Shape. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Sydney, Australia, 1–8 December 2013; pp. 1041–1048.
47. Otsu, N. Optimal Linear and Nonlinear Solutions for Least-square Discriminant Feature Extraction. In Proceedings of the 6th International Conference on Pattern Recognition (ICPR), Munich, Germany, 19–22 October 1982; pp. 557–560.
48. Liu, K.; Cheng, Y.; Yang, J. Algebraic feature extraction for image recognition based on an optimal discriminant criterion. *Pattern Recognit.* **1993**, *26*, 903–911. [CrossRef]
49. Xu, D.; Yan, S.; Zhang, L.; Zhangh, H.J.; Shum, H.Y. Concurrent Subspaces Analysis. In Proceedings of the International Conference Computer Vision and Pattern Recognition (CVPR), San Diego, CA, USA, 20–25 June 2005; pp. 203–208.
50. Martin-Felez, R.; Xiang, T. Uncooperative gait recognition by learning to rank. *Pattern Recognit.* **2014**, *47*, 379–3806. [CrossRef]
51. Iwama, H.; Okumura, M.; Makihara, Y.; Yagi, Y. The OU-ISIR Gait Database Comprising the Large Population Dataset and Performance Evaluation of Gait Recognition. *IEEE Trans. Inf. Forensics Secur.* **2012**, *7*, 1511–1521. [CrossRef]
52. Chen, X.; Weng, J.; Lu, W.; Xu, J. Multi-gait Recognition based on Attribute Discovery. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 1697–1710. [CrossRef] [PubMed]
53. Alotaibi, M.; Mahmood, A. Improved Gait recognition based on specialized deep convolutional neural networks. In Proceedings of the IEEE Applied Imagery Pattern Recognition Workshop (AIPR), Washington, DC, USA, 13–15 October 2015; pp. 1–7.
54. Kusakunniran, W. Recognizing Gaits on Spatio-Temporal Feature Domain. *IEEE Trans. Inf. Forensics Secur.* **2014**, *9*, 1416–1423. [CrossRef]
55. Xu, C.; Makihara, Y.; Li, X.; Yagi, Y.; Lu, J. Speed Invariance vs. Stability: Cross-Speed Gait Recognition Using Single-Support Gait Energy Image. In Proceedings of the 13th Asian Conference on Computer Vision (ACCV), Taipei, Taiwan, 20–24 November 2016; pp. 52–67.
56. Makihara, Y.; Tsuji, A.; Yagi, Y. Silhouette Transformation based on Walking Speed for Gait Identification. In Proceedings of the 23rd IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, CA, USA, 13–18 June 2010; pp. 717–722.
57. Kusakunniran, W.; Wu, Q.; Zhang, J.; Li, H. Gait Recognition Across Various Walking Speeds Using Higher Order Shape Configuration Based on a Differential Composition Model. *IEEE Trans. Syst. Man Cybern. Part B* **2012**, *42*, 1654–1668. [CrossRef] [PubMed]
58. Guan, Y.; Li, C.T. A robust speed-invariant gait recognition system for walker and runner identification. In Proceedings of the 6th IAPR International Conference on Biometrics (ICB), Madrid, Spain, 4–7 June 2013; pp. 1–8.
59. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R.B. Mask R-CNN. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017.