Joint Intensity and Spatial Metric Learning for Robust Gait Recognition

Yasushi Makihara[1], Atsuyuki Suzuki[1], Daigo Muramatsu[1], Xiang Li[2,1], Yasushi Yagi[1]
[1]Osaka University, [2]Nanjing University of Science and Technology
{makihara, a-suzuki, muramatsu, li, yagi}@am.sanken.osaka-u.ac.jp

Abstract

This paper describes a joint intensity metric learning method to improve the robustness of gait recognition with silhouette-based descriptors such as gait energy images. Because existing methods often use the difference of image intensities between a matching pair (e.g., the absolute difference of gait energies for the l_1-norm) to measure a dissimilarity, large intrasubject differences derived from covariate conditions (e.g., large gait energies caused by carried objects vs. small gait energies caused by the background), may wash out subtle intersubject differences (e.g., the difference of middle-level gait energies derived from motion differences). We therefore introduce a metric on joint intensity to mitigate the large intrasubject differences as well as leverage the subtle intersubject differences. More specifically, we formulate the joint intensity and spatial metric learning in a unified framework and alternately optimize it by linear or ranking support vector machines. Experiments using the OU-ISIR treadmill data set B with the largest clothing variation and large population data set with bag, β version containing carrying status in the wild demonstrate the effectiveness of the proposed method.

1. Introduction

Gait [55] is a behavioral biometric that has advantages over other biometrics such as the face, irises, or finger veins because (i) gait is available even when the subject is at a distance from a camera because it can be recognized from a relatively low-resolution image sequence [52], and (ii) a gait feature can be obtained without subject cooperation because people unconsciously exhibit their own walking styles in general. Because of these advantages, gait recognition is suitable for many potential applications such as surveillance, forensics, and criminal investigation [8, 21, 44]).

Approaches to gait recognition mainly fall into two families: model-based approaches [2,7,10,35,68,69,76,80] and model-free (appearance-based) approaches [5, 6, 16, 27, 47, 54, 60, 70, 71]. The model-based approaches fit articulated human body models to images and extract kinematic features such as joint angle sequences. While model-based approaches have robustness to some covariates such as clothing, carrying status, and partial occlusion, they require a relatively high image resolution to get reasonable human model fitting results and incur high computational costs.

Figure 1. Examples of carrying status. Carried objects (indicated by yellow dotted circles) appear at various positions and hence spatial metric learning does not work for robust gait recognition.

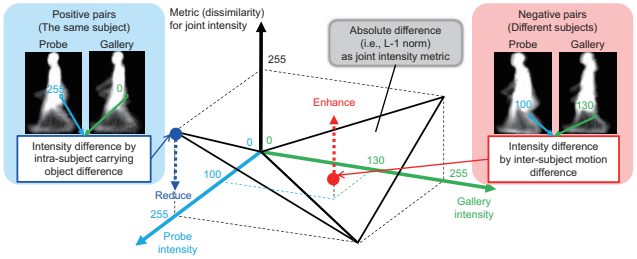


Figure 2. Concept of joint intensity metric learning. A conventional joint intensity metric (i.e., the absolute difference of the l_1-norm in this figure) returns a large dissimilarity for intrasubject carried object differences (e.g., intensity level 255 vs. 0), while it returns a small dissimilarity for intersubject motion difference (e.g., intensity level 100 vs. 130) in GEI. We therefore want to reduce the dissimilarity of the joint intensities derived from intrasubject differences while enhancing the dissimilarity for those derived from the intersubject differences by joint intensity metric learning.

The appearance-based approaches directly use input or silhouette images in a holistic way to extract gait features without model fitting, and hence they generally work well, even for relatively low-resolution images, and incur low computational costs. In particular, silhouette-based representations such as gait energy images (GEIs) [16], frequency-domain features (FDFs) [47], chrono-gait images [70], and Gabor GEIs [66], are dominant in the gait recognition community because of their simple yet effective properties. The appearance-based approaches, however, often suffer from large intrasubject appearance changes due to covariates such as clothing [5, 6, 20, 36, 58], carrying status [11, 65, 67], view [22, 28, 30, 32, 43, 47, 61, 62, 73], and

walking speed [1, 13, 29, 31, 41, 48, 49, 64].

The most popular way to gain robustness to covariates is to incorporate spatial metric learning such as linear discriminant analysis (LDA) [16, 40, 70], general tensor discriminant analysis (GTDA) [65, 66], discriminant analysis with tensor representation (DATER) [74, 77], or the random subspace method (RSM) [14, 15]. It is, however, difficult to cover all the variations only by spatial metric learning because the spatial positions affected by covariates such as clothing and carrying status are quite different depending on the instances (e.g., backpacks, shoulder bags, handbags, and briefcases), as shown in Fig. 1.

However, taking a closer look at an intensity metric aspect for gait recognition, we note that the difference of intensities between a matching pair (e.g., the absolute difference of gait energies for the l_1-norm) is often used to measure a dissimilarity, where a larger difference of intensities naturally results in a larger dissimilarity. Therefore, large intrasubject differences derived from covariate conditions, e.g., the difference between high intensity (large gait energies in a GEI representation) from foreground carried objects and low intensity (small gait energies) from the background, may overwhelm subtle intersubject differences, e.g., the differences of medium intensities (middle-level gait energies) derived from motion differences, as shown in Fig. 2.

A few studies on gait recognition focus on this intensity metric aspect. Bashir et al. [5] proposed gait entropy image (GEnI), which transforms a gait energy into a Shannon entropy for each pixel to enhance the dynamic parts while neglecting static parts to gain robustness to clothing and carrying status (see Fig. 4(b) for an example). Moreover, a masked GEI was proposed in [6], where the static parts of the GEI (i.e., small entropy regions) are masked out. Both methods transform the complete background (the minimum gait energy) and complete foreground (the maximum gait energy) to zero, and consequently make their difference zero, which is an excessive suppression of intrasubject variation that risks overwhelming the intersubject difference derived from individual shape variation.

We therefore introduce a metric on joint intensity to mitigate the large intrasubject differences and leverage the subtle intersubject differences, as shown in Fig. 2. We learn such a metric so as to separate pairs of the same subjects and different subjects well in a data-driven way. The contributions of this work are three-fold.

1. A data-driven joint intensity metric

While the existing intensity-oriented methods such as GEnI [5] and masked GEI [6] are designed in a handcrafted way and also completely discard individual shape differences, the proposed method learns the joint intensity metric in a data-driven way. More specifically, we learn the joint intensity metric to realize a good tradeoff between suppressing intrasubject differences and enhancing intersubject differences using a training set that includes covariate conditions such as clothing and carrying status.

2. A unified framework for joint intensity and spatial metric learning

We optimize not only the joint intensity metric but also a spatial metric (weighting on each spatial position) in a unified framework. More specifically, we define a dissimilarity measure for a matching pair as a bilinear form of a joint intensity metric vector and a spatial metric vector, and then define an objective function that maximizes a margin between positive (the same subjects) and negative (different subjects) pairs. Consequently, we can alternately optimize the joint intensity metric and spatial metric in a framework of a linear support vector machine (SVM) or ranking SVM.

3. State-of-the-art accuracy on robust gait recognition

We achieved state-of-the-art accuracy on gait recognition under clothing and carrying status variations using publicly available gait databases containing the largest clothes variations (up to 32 types) and carrying status in the wild.

2. Related work

Joint intensity histogram

Joint intensity histograms were originally defined for a pair of spatially adjacent pixels and applied for various purposes such as texture analysis [17] and template matching [18]. Thereafter, some studies defined the joint intensity at the same spatial position for an image pair in the same way as the proposed method. For example, Kita et al. [23–25] constructed a joint intensity histogram between a background image (sequence) and an input image, and then detected the changed region (foreground) by finding a background cluster from the joint intensity histogram, which improves the robustness to illumination change compared with simple background subtraction. Moreover, the registration of 3D volume data is performed using joint intensity histogram in medical image analysis [37].

Although the joint intensity histogram has been widely used in a various research fields as above, there is no study on joint intensity metric learning for better recognition of individuals, to the best of our knowledge.

Joint probability modeling for sequence matching

The proposed joint intensity metric learning for image pair matching, is conceptually similar to joint probability modeling for a family of sequence matching or alignment (e.g., approximate string matching [4] and DNA sequence alignment [53]) methods to some degree. In short, a joint probability for a pair of unit entries (e.g., characters for approximate string matching and amino acids for DNA sequence alignment) can be used to define the costs of insertion, deletion, and substitution of the unit entries for an elastic sequence matching algorithm such as dynamic programming. Because the target of this work is an image, the

joint intensity metric needs to be trained in conjunction with a spatial metric for better performance in total. In contrast, sequence matching does not consider such a spatial metric.

3. Joint intensity and spatial metric learning

3.1. Observation on spatial metric learning

To clarify the difference between the proposed method and conventional spatial metric learning, we first review typical spatial metric learning. Given a gray-scale image with height H and width W (i.e., $N_S = WH$ pixels) whose image intensity at position (x, y) is denoted as $v_{(x,y)} \in \mathbb{Z}$, an unfolded image intensity vector is denoted as $\boldsymbol{v} = [v_{(1,1)}, \ldots, v_{(W,1)}, \ldots, v_{(1,H)}, \ldots, v_{(W,H)}]^T = [v_1, \ldots, v_{N_S}]^T \in \mathbb{Z}^{N_S}$.[1]

Thereafter, given a matching pair of images, i.e., a probe (or a query) $\boldsymbol{v}^P$ and a gallery (or an enrollment) $\boldsymbol{v}^G$, their l_1-norm dissimilarity is

$$D(\boldsymbol{v}^P, \boldsymbol{v}^G) = \sum_{i=1}^{N_S} |v_i^P - v_i^G|. \tag{1}$$

Moreover, for better classification, we often incorporate spatial weights into the dissimilarity measure as

$$D(\boldsymbol{v}^P, \boldsymbol{v}^G; \boldsymbol{w}_S) = \sum_{i=1}^{N_S} w_{S,i} |v_i^P - v_i^G|, \tag{2}$$

where $\boldsymbol{w}_S = [w_{S,1}, \ldots, w_{S,N_S}]^T$ is a spatial weight vector. The above spatially weighted l_1-norm is adopted in a linear SVM framework for example, where a matching pair is classified into two classes: a genuine pair (or the same subject, positive sample) of an imposter pair (or different subjects, negative sample) [51].

Another popular metric is the Mahalanobis distance

$$D(\boldsymbol{v}^P, \boldsymbol{v}^G; M) = (\boldsymbol{v}^P - \boldsymbol{v}^G)^T M (\boldsymbol{v}^P - \boldsymbol{v}^G), \tag{3}$$

where $M \in \mathbb{R}^{N_S \times N_S}$ is a semi-positive definite matrix. Various metric learning approaches fall into this category, for example, ranging from classical projection-based approaches such as LDA [57], 2D-LDA [39], GTDA [66] and concurrent subspace analysis (CSA) [75], DATER [77] to recent metric learning-based approaches such as rank-based distance metric learning [34], large margin nearest neighbor (LMNN) [72], probabilistic relative distance comparison (PRDC) [81], and random ensemble metrics (REMetric) [26], which are often used in studies on gait recognition [6, 16, 41, 74] and person re-identification.

We note that intensities appear in a subtraction form $(\boldsymbol{v}^P - \boldsymbol{v}^G)$ and hence the original values do not affect the dissimilarity, e.g., $(v_i^P, v_i^G) = (100, 0)$ and $(v_i^P, v_i^G) = (200, 100)$ result in the same dissimilarity. In addition, as absolute difference $|\boldsymbol{v}^P - \boldsymbol{v}^G|$ increases, a dissimilarity monotonically increases linearly (Eq. (2)) or quadratically (Eq. (3)), and hence a dissimilarity for $(v_i^P, v_i^G) = (255, 0)$ is larger than $(v_i^P, v_i^G) = (100, 130)$.

The above properties are generally reasonable and hence exploited in most image matching algorithms. It is, however, not true under a certain situation. Consider gait recognition using a silhouette-based representation such as GEI [16] under covariate variations such as clothing and carrying status. As shown in Fig. 2, an intrasubject carried object difference (e.g., $(v_i^P, v_i^G) = (255, 0)$) returns a large dissimilarity, while an intersubject motion difference (e.g., $(v_i^P, v_i^G) = (100, 130)$) returns a small dissimilarity. Therefore, if we learn an appropriate dissimilarity metric on joint intensities to reduce intrasubject difference while enhancing the intersubject difference, we can improve the robustness against covariate conditions.

Although the l_p-norm [33, 56] ($0 < p < 1$) may realize this to some extent (i.e., small absolute differences are somewhat more leveraged and large absolute differences are somewhat reduced relative to the l_1-norm), a more flexible operation, such as setting a larger dissimilarity for smaller absolute differences and setting a different dissimilarity for the same absolute differences, is still impossible.

3.2. Representation of dissimilarity measure

Based on the above observation, we extend the spatially weighted l_1-norm (Eq. (2)) so that it represents a more general metric for joint intensity as follows:

$$D(\boldsymbol{v}^P, \boldsymbol{v}^G; \boldsymbol{w}_S, \eta) = \sum_{i=1}^{N_S} w_{S,i} \eta(v_i^P, v_i^G), \tag{4}$$

where $\eta(p, g)$ is a spatially independent dissimilarity metric for joint intensity (p, g) (e.g., $\eta(p, g) = |p - g|$ for l_1-norm).

In other words, $\eta(p, g)$ is regarded as a mapping function from an intensity pair (p, q) to a dissimilarity. Function $\eta(p, g)$ takes as arguments a set of intensity pairs $\{(p, g) | p \in \mathcal{S}_I, g \in \mathcal{S}_I\}$, where $\mathcal{S}_I = \{0, \ldots, I_{\max}\}$ is a set of integer intensity levels from 0 (the minimum) to $I_{\max}$ (the maximum. Hence, we may represent mapping function $\eta(p, g)$ using a set of $N_I (= (I_{\max} + 1)^2)$ combinations of dissimilarities $\{w_{I,(p,g)}\} (p = 0, \ldots, I_{\max}, g = 0, \ldots, I_{\max})$ as

$$\eta(p, g) = w_{I,(p,g)} = \sum_{k=0}^{I_{\max}} \sum_{l=0}^{I_{\max}} \delta_{k,p} \delta_{l,g} w_{I,(k,l)}, \tag{5}$$

where $\delta_{i,j}$ is Kronecker's delta.

We further reformulate the above equation in a simple inner product form as

$$\eta(p, g) = \boldsymbol{x}_{(p,g)}^T \boldsymbol{w}_I, \tag{6}$$

[1] The positional subscription-based notation $v_{(x,y)}$ and single sequential index-based notation v_i are used interchangeably in this paper.

where $\boldsymbol{w}_I=[w_{I,(0,0)},\ldots,w_{I,(0,I_{\max})},\ldots,w_{I,(I_{\max},0)},\ldots,w_{I,(I_{\max},I_{\max})}]^T = [w_{I,1},\ldots,w_{I,N_I}]^T \in \mathbb{R}^{N_I}$[2] is a joint intensity metric vector, and $\boldsymbol{x}_{(p,g)} = [x_{(p,g),1},\ldots,x_{(p,g),N_I}]^T$ is an N_I-dimensional indicator vector whose $(p(I_{\max}+1)+g)$-th component is activated $(=1)$ while the other components are zero-padded. Specifically, the indicator vector is derived from Eqs. (5) and (6) as

$$\boldsymbol{x}_{(p,g)}=[\delta_{p,0}\delta_{g,0},\ldots,\delta_{p,0}\delta_{g,I_{\max}},\ldots,\delta_{p,I_{\max}}\delta_{g,0},\ldots,\delta_{p,I_{\max}}\delta_{g,I_{\max}}]^T. \quad (7)$$

Finally, we can reformulate the whole dissimilarity measure (Eq. (4)) as a bilinear form of spatial metric vector $\boldsymbol{w}_S$ and joint intensity metric vector $\boldsymbol{w}_I$ as

$$\begin{aligned} D(\boldsymbol{v}^P,\boldsymbol{v}^G;\boldsymbol{w}_S,\boldsymbol{w}_I) &= \sum_{i=1}^{N_S}\sum_{j=1}^{N_I} w_{S,i}w_{I,j}x_{(v_i^P,v_i^G)} \\ &= \boldsymbol{w}_S^T X_{(\boldsymbol{v}^P,\boldsymbol{v}^G)}\boldsymbol{w}_I, \end{aligned} \quad (8)$$

where $X_{(\boldsymbol{v}^P,\boldsymbol{v}^G)}=[\boldsymbol{x}_{(v_1^P,v_1^G)},\ldots,\boldsymbol{x}_{(v_{N_S}^P,v_{N_S}^G)}]^T \in \{0,1\}^{N_S\times N_I}$ is an indicator matrix.

3.3. Metric learning

We next consider learning joint intensity metric $\boldsymbol{w}_I$ and spatial metric $\boldsymbol{w}_S$ in the dissimilarity measure (Eq. (8)) so as to well separate the genuine and imposter pairs in a data-driven way. More specifically, a training set is given as a set $\mathcal{S}=\{(\boldsymbol{v}^P,\boldsymbol{v}^G,t)\}$ of image pairs $\boldsymbol{v}^P$, $\boldsymbol{v}^G$, and a corresponding class indicator $t\in\{-1,1\}$, where 1 and -1 indicate genuine and imposter pairs, respectively.

Similar to prior work on spatial metric learning, we exploit an SVM framework and then introduce the following objective function to be minimized:

$$\begin{aligned} J(\boldsymbol{w}_S,\boldsymbol{w}_I) &= \frac{1}{2}\|\boldsymbol{w}_S\|^2+\frac{1}{2}\|\boldsymbol{w}_I\|^2 \\ &+ C\sum_{(\boldsymbol{v}^P,\boldsymbol{v}^G,t)\in\mathcal{S}} l(\boldsymbol{w}_S^T X_{(\boldsymbol{v}^P,\boldsymbol{v}^G)}\boldsymbol{w}_I+b,t), \end{aligned} \quad (9)$$

where the first two terms are regularization terms for margin maximization for spatial metric $\boldsymbol{w}_S$ and joint intensity metric $\boldsymbol{w}_I$, and the third term is a data term, i.e., a penalty for a soft margin, which contains bias parameter b for optimization in SVM and hyper-parameter C to balance the regularization and data terms. Function $l(y,t)$ is a hinge loss function, defined as

$$l(y,t)=w_t\max(0,1-ty), \quad (10)$$

where w_t is a class specific weight that we define as $w_1=1$ and $w_{-1}=N_{gen}/N_{imp}$ so as to balance the genuine and imposter samples, where N_{gen} and N_{imp} are the numbers of genuine and imposter training samples, respectively.

Because the objective function (Eq. (9)) has a bilinear data term, we adopt an alternate optimization framework. Once we fix spatial metric $\boldsymbol{w}_S$, the objective function is

$$J_I(\boldsymbol{w}_I)=\frac{1}{2}\|\boldsymbol{w}_I\|^2+C\sum_{(\boldsymbol{v}^P,\boldsymbol{v}^G,t)\in\mathcal{S}} l(\boldsymbol{h}_{(\boldsymbol{v}^P,\boldsymbol{v}^G)}^T\boldsymbol{w}_I+b,t), \quad (11)$$

$$\boldsymbol{h}_{(\boldsymbol{v}^P,\boldsymbol{v}^G)}=X_{(\boldsymbol{v}^P,\boldsymbol{v}^G)}^T\boldsymbol{w}_S. \quad (12)$$

Note that vector $\boldsymbol{h}_{(\boldsymbol{v}^P,\boldsymbol{v}^G)}\in\mathbb{R}^{N_I}$ is regarded as a sort of joint intensity histogram (see Fig. 3(d)), as its $(p(I_{\max}+1)+g)$-th component is derived from Eqs. (7) and (8) as

$$\boldsymbol{h}_{(\boldsymbol{v}^P,\boldsymbol{v}^G),(p,g)}=\sum_{i=1}^{N_S} w_{S,i}\delta_{v_i^P,p}\delta_{v_i^G,g}, \quad (13)$$

which is equivalent to counting the number of pixels whose joint intensity coincides with (p,g).

In contrast, once we fix joint intensity metric $\boldsymbol{w}_I$, the objective function is

$$J_S(\boldsymbol{w}_S)=\frac{1}{2}\|\boldsymbol{w}_S\|^2+C\sum_{(\boldsymbol{v}^P,\boldsymbol{v}^G,t)\in\mathcal{S}} l(\boldsymbol{m}_{(\boldsymbol{v}^P,\boldsymbol{v}^G)}^T\boldsymbol{w}_S+b,t), \quad (14)$$

$$\boldsymbol{m}_{(\boldsymbol{v}^P,\boldsymbol{v}^G)}=\boldsymbol{w}_I^T X_{(\boldsymbol{v}^P,\boldsymbol{v}^G)}^T. \quad (15)$$

Note that vector $\boldsymbol{m}_{(\boldsymbol{v}^P,\boldsymbol{v}^G)}\in\mathbb{R}^{N_S}$ indicates a set of dissimilarities at individual spatial positions (later called "spatial dissimilarity," see Fig. 3(e)) and that its i-th component is simply derived from Eqs. (4) and (5) as $m_{(\boldsymbol{v}^P,\boldsymbol{v}^G),i}=w_{I,(v_i^P,v_i^G)}$.

Each objective function (Eqs. (11) and (14)) is efficiently minimized by the linear SVM framework, and hence we alternately optimize the joint intensity metric $\boldsymbol{w}_I$ and spatial metric $\boldsymbol{w}_S$ given a training set (Fig. 3(a)). Specifically, we (i) compute joint intensity histogram $\boldsymbol{h}_{(\boldsymbol{v}^P,\boldsymbol{v}^G)}$ (Fig. 3(d)) under fixed spatial metric $\boldsymbol{w}_S$ (Fig. 3(c)) and update joint intensity metric $\boldsymbol{w}_I$ (Fig. 3(b)) by minimizing Eq. (11) and (ii) compute spatial dissimilarity $\boldsymbol{m}_{(\boldsymbol{v}^P,\boldsymbol{v}^G)}$ (Fig. 3 (e)) under fixed intensity metric $\boldsymbol{w}_I$ (Fig. 3 (b)) and update spatial metric $\boldsymbol{w}_S$ (Fig. 3 (c)) by minimizing Eq. (14). For this alternate optimization framework, we initialize joint intensity metric $\boldsymbol{w}_I$ by the absolute difference, i.e., $w_{I,(p,g)}=|p-g|\ \forall(p,g)$, while we initialize the spatial metric to a uniform weight, i.e., $w_{S,i}=1\ \forall i$.

3.4. Extension to identification mode

Because the objective function (Eq. (9)) introduced in the previous section aims to separate genuine and imposter pairs based only on the dissimilarity of the pair, it is suitable for the verification mode, i.e., the one-to-one matching used in, for example, access control. In contrast, when considering an identification mode, i.e., finding the top n matched

[2]The joint intensity subscription-based notation $\{w_{I,(p,q)}\}(p\in\mathcal{S}_I, g\in\mathcal{S}_I)$ and a single sequential index-based notation $\{w_{I,j}\}(j=1,\ldots,N_I)$ are used interchangeably in this paper.

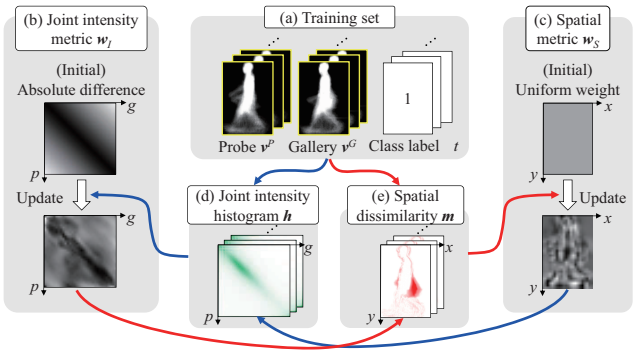


Figure 3. Framework of joint intensity and spatial metric learning. Given a training set (a), the joint intensity metric (b) is updated using joint intensity histograms (d) under a fixed spatial metric (c), while the spatial metric (c) is updated using spatial dissimilarity (e) under a fixed joint intensity metric (b).

subjects from a gallery list (e.g., person re-identification), it is not the dissimilarity of a specific pair but a relative dissimilarity between the genuine and imposter galleries for a specific probe (query) that is important. For this purpose, it is well known that a triplet loss function works well [50,51], where the optimizer drives a dissimilarity between probe $\boldsymbol{v}^P$ and an imposter gallery $\boldsymbol{v}^{G_{imp}}$ to be larger than a dissimilarity between the same probe $\boldsymbol{v}^P$ and genuine gallery $\boldsymbol{v}^{G_{gen}}$. Specifically, the objective function for the triplet loss is

$$J(\boldsymbol{w}_S, \boldsymbol{w}_I) = \frac{1}{2}\|\boldsymbol{w}_S\|^2 + \frac{1}{2}\|\boldsymbol{w}_I\|^2 + C\sum_{(\boldsymbol{v}^P, \boldsymbol{v}^{G_{gen}}, \boldsymbol{v}^{G_{imp}}) \in \mathcal{S}_{trip}} l(\boldsymbol{w}_S^T(X_{(\boldsymbol{v}^P, \boldsymbol{v}^{G_{imp}})} - X_{(\boldsymbol{v}^P, \boldsymbol{v}^{G_{gen}})})\boldsymbol{w}_I, 1), \quad (16)$$

where $\mathcal{S}_{trip}$ is a set of triplets of a probe, an imposter gallery, and a genuine gallery.

We again solve this problem by alternate optimization of joint intensity metric $\boldsymbol{w}_I$ and spatial metric $\boldsymbol{w}_S$ in a similar way to the framework described in the previous section. The only difference is that we used primal RankSVM [9] for this optimization instead of linear SVM.

3.5. Regularization for proximity

Because each element in spatial metric $\boldsymbol{w}_S$ is optimized regardless of the spatial proximity in the above method, it may easily over-fit the training set and hence lose its generalization capability. We therefore introduce a regularizer for the spatial proximity. Simply, we introduce the 1st-order smoothness on the adjacent points as

$$R_S(\boldsymbol{w}_S) = \lambda_S \sum_{x,y} \left\{ (w_{S,(x,y)} - w_{S,(x-1,y)})^2 + (w_{S,(x,y)} - w_{S,(x,y-1)})^2 \right\} = \frac{1}{2}\lambda_S \boldsymbol{w}_S^T \Sigma_S \boldsymbol{w}_S, \quad (17)$$

where $\Sigma_S \in \mathbb{R}^{N_S \times N_S}$ and λ_S are a symmetric matrix and a coefficient for regularizing the spatial metric, respectively.

The objective function for spatial metric $\boldsymbol{w}_S$ (Eq. (14)) is now reformulated as

$$J'_S(\boldsymbol{w}_S) = \frac{1}{2}\boldsymbol{w}_S^T \tilde{\Sigma}_S \boldsymbol{w}_S + C\sum_{(\boldsymbol{v}^P, \boldsymbol{v}^G, t) \in \mathcal{S}} l(\boldsymbol{m}_{(\boldsymbol{v}^P, \boldsymbol{v}^G)}^T \boldsymbol{w}_S + b, t), \quad (18)$$

$$\tilde{\Sigma}_S = I + \lambda_S \Sigma_S, \quad (19)$$

where $\tilde{\Sigma}_S$ is a new matrix for the spatial regularization considering both margin and spatial proximity. To solve this problem, [12] suggests changing variables as $\tilde{\boldsymbol{w}}_S = \tilde{\Sigma}_S^{1/2}\boldsymbol{w}_S$ and $\tilde{\boldsymbol{m}} = \tilde{\Sigma}_S^{-1/2}\boldsymbol{m}$ and to apply the linear SVM solver to the new variables.

Similarly, it is worth considering the proximity for the joint intensity metric. Instead of the simple horizontal and vertical proximity adopted in the spatial metric as above, we consider the proximity for diagonal and antidiagonal directions in the joint intensity by taking it into account a relation between the joint intensity and dissimilarity. Because the absolute difference of intensities is kept along the diagonal direction (e.g., (p, g) and $(p-1, g-1)$), while it changes along the antidiagonal direction (e.g., (p, g) and $(p-1, g+1)$), we set a stronger regularization on the diagonal direction than on antidiagonal direction and set a medium-strength regularization on the horizontal and vertical directions. Specifically, we introduce the following regularizer for the joint intensity:

$$R_I(\boldsymbol{w}_I) = \lambda_I \sum_{p,g} \Big\{ (w_{I,(p,g)} - w_{I,(p-1,g-1)})^2 + r(w_{I,(p,g)} - w_{I,(p-1,g+1)})^2 + \sqrt{2r}(w_{I,(p,g)} - w_{I,(p-1,g)})^2 + \sqrt{2r}(w_{I,(p,g)} - w_{I,(p,g-1)})^2 \Big\} = \frac{1}{2}\lambda_I \boldsymbol{w}_I^T \Sigma_I \boldsymbol{w}_I, \quad (20)$$

where $\Sigma_I \in \mathbb{R}^{N_I \times N_I}$ and λ_I are a symmetric matrix and a coefficient for regularizing the joint intensity metric, respectively, and $r (\ll 1)$ is a hyperparameter for controlling the strength between the diagonal and antidiagonal directions. Similarly to the spatial proximity, we can include the joint intensity proximity by changing the variables.

Moreover, we can do the same thing in a ranking SVM framework (Eq. (16)), although the details are omitted due to page limitations.

4. Experiments

4.1. Data set

We conducted our experiments using two gait databases: the OU-ISIR Gait Database, Treadmill Dataset B (OUTD-B) [46], which includes the largest clothing variations, and the OU-ISIR Gait Database, Large Population data set with bag, β version (OU-LP-Bag β)[3], which includes carrying

[3]Available at http://www.am.sanken.osaka-u.ac.jp/BiometricDB/index.html.

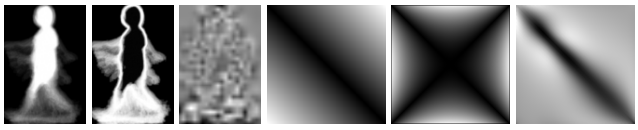
status variations in the wild. In both databases, we extract 88×128 pixel-sized GEIs as gait features.

OUTD-B includes 68 subjects with at most 32 combinations of different clothing. Walking image sequences on the treadmill were captured from a side view. The whole dataset is divided into three subsets: a training set, a gallery set, and a probe set. In the training set, there are 446 sequences of 20 subjects with a range of 15 to 28 different combinations of clothing. The gallery set and probe set form the testing set, and are composed of 48 subjects, that were disjoint from the 20 subjects in the training set. The gallery contains only standard clothing types (i.e., regular pants and full shirts), while the probe set includes 856 sequences of other remaining clothing types.

As for carrying status, because the existing gait databases such as SOTON Large Database [63], USF HumanID [60], CASIA Gait Database B [78], and TUM GAID [19] contain only a few variations, we utilized gait database that includes various carrying statuses. More specifically, gait data in this database were collected in conjunction with an experience-based demonstration of video-based gait analysis at a science museum. While each subject enjoyed the demonstration, he or she was asked to walk a path twice: once carrying his or her objects and once without carrying objects, after agreeing to an informed consent for the research purpose use of the gait data (see [45] for details). As a result, two walking image sequences of 2,070 subjects with various carrying statuses were collected. Similarly to OUTD-B, the whole dataset is 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 1,036 subjects that is disjoint from the 1,034 subjects in the training set. The gallery contains a gait image sequence without carried objects, while the probe set includes gaits with carried objects. Because the number of imposter pairs in the training set is quite large (1,000,000+), we randomly chose 400,000 imposter pairs for more efficient computation for training. In addition, unlike OUTD-B, this dataset contains a single period of gait features for each sequence.

4.2. Parameter settings

The proposed method contains some hyper-parameters. We experimentally set the default hyperparameters for the metric learning as $C = 1.0$, $\lambda_S = \lambda_I = 0.2$, and $r = 0.1$. In addition, for more efficient computation for training, we downsampled joint intensity metric $\boldsymbol{w}_I$ and spatial metric $\boldsymbol{w}_S$. Specifically, we reduced the dimensions of joint intensity metric $\boldsymbol{w}_S$ from 256×256 to 32×32, while we reduced the dimension of spatial metric $\boldsymbol{w}_S$ from 88×128 to 22×32. We refer the reader to the supplementary material for a sensitivity analysis of these hyperparameters.



(a) (b) (c) (d) (e) (f)

Figure 4. Learnt metrics of the joint intensity and spatial metrics for the OU-LP-Bag β data set. (a) Sample GEI, (b) sample GEnI, (c) learnt spatial metric using linear SVM under the absolute-difference joint intensity metric, (d) absolute difference as a joint intensity metric for the l_1-norm (i.e., w/o metric learning), (e) joint intensity metric for GEnI (i.e., hand-crafted), and (f) learnt joint intensity metric using linear SVM under uniform spatial weight. Brighter values indicate larger weights in (c) and dissimilarity in (d–f), respectively. The metrics shown in (c) and (f) are enlarged to the original size by linear interpolation from the downsampled learnt ones.

4.3. Learnt metrics

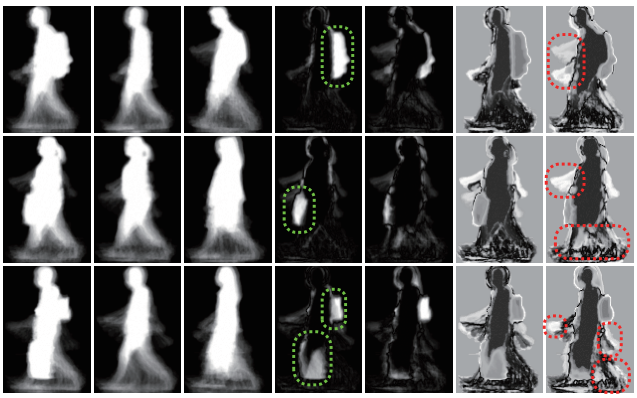
In this section, we analyze the learnt metrics and show typical matching examples using the learnt metrics. Figure 4 shows the learnt joint intensity and spatial metrics as well as GEI [16] and GEnI [5] samples.

The learnt spatial metrics were often successfully interpreted in prior work [51, 79] (e.g., a region around the silhouette contour is enhanced, while a region affected by carried objects is reduced), because they handle only a limited variations of clothing or carrying status. The spatial metric (Fig. 4(c)) learnt with diverse carrying conditions (Fig. 1), is, however, too complex to be interpreted, which implies the difficulty or over-fitting risk when we use the spatial metric alone.

In contrast, the learnt joint intensity metric (Fig. 4(f)) is more easily interpreted. Compared with the joint intensity metrics for absolute-difference (Fig. 4(d)) and for GEnI [5] (Fig. 4(e)), the learnt joint intensity metric (Fig. 4(f)) has some interesting points. First, while small dissimilarities are assigned to regions close to the diagonal line, i.e., subtle motion differences appeared in GEI in Figs. 4(d) and (e), relatively large dissimilarities are assigned to such regions for the learnt metric (Fig. 4(f)), which indicates that the learnt metric successfully leverages the informative intersubject motion difference.

Moreover, for intensity pairs of complete foreground and background regions (i.e., near the top-right and bottom-left corners), which is partly derived from the intrasubject carrying status variations, the absolute difference returns the largest dissimilarity (Fig. 4(d)). In contrast, because GEnI [5] enhances motion regions while discarding static regions regardless of the background or foreground, such intensity pairs return the smallest dissimilarity (Fig. 4(e)). Generally speaking, GEnI cannot distinguish two intensity values in GEI: v and $(255 - v)$, i.e., an anti-diagonal line in Fig. 4(e), this hand-crafted joint intensity metric may overwhelm the important intersubject variations.

Unlike these two extreme joint intensity metrics, the pro-



(a) (b) (c) (d) (e) (f) (g)

Figure 5. Matching examples for three probes with different types of carried objects. (a) Probe. (b) Gallery (genuine). (c) Gallery (imposter). (d) and (e) Spatial dissimilarity with absolute difference for a true match pair ((a) and (b)) and false match pair ((a) and (c)), respectively. (f) and (g) Spatial dissimilarity with the learnt joint intensity metric shown in Fig. 4(f) for a true match pair ((a) and (b)) and false match pair ((a) and (c)), respectively. In (d)–(g), darker and brighter regions indicate similar and dissimilar regions, respectively. While the dissimilarities with the absolute difference for the true match pairs are larger than those for the false match pairs because of the carried objects (green-dotted circled regions), they are suppressed to some extent with the learnt joint intensity metric in (f). In contrast, the motion difference region between the false match pairs are enhanced by the learnt joint intensity metric in (g), and hence the dissimilarities using the learnt joint intensity metric for the false match pairs are larger than those for the true match pairs.

posed joint intensity metric (Fig. 4(f)) learns an appropriate dissimilarity for the intensity pairs of complete foreground and background regions in a data-driven way, i.e., neither the maximum dissimilarity, as in Fig. 4(d), nor completely zero, as in Fig. 4(e), but moderate value which is smaller than dissimilarity for intersubject motion variations around the diagonal line, as shown in Fig. 4(f).

Thanks to the properties of the learnt joint intensity metric, we can successfully recover the failure modes of l_1 norm-based gait recognition under various carrying conditions (e.g., a subject carrying a backpack or briefcase) as shown in Fig. 5. In fact, we can see that the learnt joint intensity mitigates the effect of carrying status regardless of its spatial position, while leveraging intersubject motion differences, which results in successful recognition.

4.4. Analysis of individual modules

In this section, we analyze the effect of individual components of the proposed metric learning (ML). We consider l_1-norm (Eq. (1)) (denoted as w/o ML) as a baseline, and also three variants as benchmarks: (1) spatial metric learning only (S-ML), (2) joint intensity metric learning only (JI-ML), and (3) both joint intensity and spatial metric learning (JIS-ML). In addition, we train the above three metrics by both linear SVM and ranking SVM. We evaluated them in both verification (one-to-one matching) and identification modes for both the test and training sets of OUTD-B, and report in Table 1 the equal error rate (EER) of the false rejection rate (FRR) of the genuine pairs and the false acceptance rate (FAR) of the imposter pairs for the verification mode. We also report the rank-1 identification rate in the identification mode.

Table 1. EER [%] and rank-1 identification rate (denoted as Rank-1) [%] for individual metrics and solvers for OUTD-B. Bold and italic bold indicate the best and the second best performances, respectively.

Set	Test		Training	
Method	EER	Rank-1	EER	Rank-1
w/o ML	26.3	61.6	27.2	72.3
S-ML (Linear SVM)	27.2	12.0	7.0	84.5
JI-ML (Linear SVM)	13.9	67.9	14.6	79.1
JIS-ML (Linear SVM)	14.5	56.4	**1.5**	96.9
S-ML (Ranking SVM)	14.6	60.2	7.9	***97.7***
JI-ML (Ranking SVM)	***13.4***	***71.5***	11.3	91.3
JIS-ML (Ranking SVM)	**11.0**	**74.5**	***3.5***	**99.8**

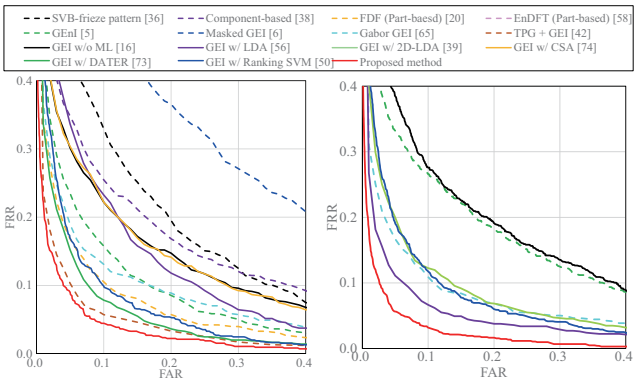
As a result, JIS-ML (Ranking SVM) yielded the best or second best accuracies, and JIS-ML (Linear SVM) yielded the best accuracy for the training set in verification mode (i.e., the lowest EER), which is reasonable when taking the properties of linear SVM and ranking SVM into account. Another interesting aspect is that S-ML yielded better accuracy than JI-ML for the training set while S-ML is performed worse than JI-ML on the test set. This indicates that the spatial metric tends to over-fit to the covariate conditions in the training set (i.e., the spatial positions are affected by the clothing variations of training subjects) and hence suffers from generalization errors. In contrast, the joint intensity metric successfully maintains its accuracy (e.g., the EERs of the test set are not significantly degraded from those of the training set,[4]) and hence has better generalization capability than the spatial metric.

4.5. Comparison with the state-of-the-arts

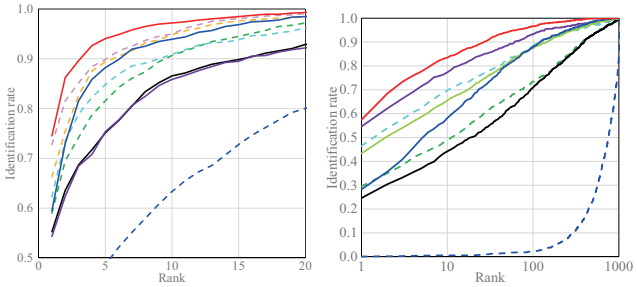
In this section, we compare the proposed method of joint intensity and spatial metric learning using ranking SVM (JIS-ML (Ranking SVM)) with state-of-the-art benchmarks. To evaluate the verification mode, we applied probe-dependent z-normalization [3], i.e., we compute the dissimilarity scores between a specific probe and all the galleries, and normalize them so as that their means and standard deviations are 0 and 1, respectively, because the state-of-the-art [5, 6, 16, 20, 36, 38, 42, 74, 75] are also evaluated with z-normalization for OUTD-B.

As for the gait features used, in addition to GEI, which is the most popular descriptor in the gait recognition research

[4]Note that rank-1 identification rates depend on the gallery size (20 and 48 subjects for training and test sets, respectively).



(a) ROC curves with z-normalization



(b) CMC curves

Figure 6. ROC and CMC curves for the comparison experiments (left: OUTD-B, right: OU-LP-Bag β). Note that some benchmarks do not provide curves.

Table 2. Comparison experiments. Rank-1 and z-EER indicate rank-1 identification rate [%] and EER with z-normalization [%]. Bold and italic bold indicate the best and second best accuracies. N/A and "-" indicate not applicable and not provided, respectively.

Dataset	OUTD-B		OU-LP-Bag β	
Method	z-EER	Rank-1	z-EER	Rank-1
SVB-frieze pattern [36]	19.81	-	-	-
Component-based [38]	18.25	-	-	-
FDF (Part-based) [20]	10.26	66.3	-	-
EnDFT (Part-based) [59]	-	72.8	-	-
GEnI [5]	12.81	59.0	18.82	29.5
Masked GEI [6]	28.15	28.0	61.95	0.1
Gabor GEI [66]	11.80	62.3	10.48	46.4
Gabor GEI w/ RSM [15]	N/A	**90.7**	N/A	N/A
TPG + GEI [42]	***7.10***	-	-	-
GEI w/o ML [16]	16.21	55.3	19.59	24.6
GEI w/ LDA [57]	15.63	54.3	8.10	***54.6***
GEI w/ 2DLDA [39]	8.91	70.7	11.47	43.3
GEI w/ CSA [75]	16.00	-	-	-
GEI w/ DATER [74]	8.72	-	-	-
GEI w/ Ranking SVM [51]	10.75	58.4	10.81	28.3
Proposed method	**6.66**	***74.5***	**5.45**	**57.4**

community, we also chose gait features that are robust against clothing variations and/or carrying status: shape variation-based (SVB) frieze pattern [36], component-based features [38], part-based FDF [20], part-based ENtropy of the Discrete Fourier Transform (EnDFT) [59], GEnI [5], Masked-GEI [6], Gabor GEI [5], Gabor GEI with a random subspace method (RSM) [15], and two-point gait (TPG) + GEI [42]. Moreover, we applied a variety of spatial metric learning methods to GEI: LDA [57], 2D-LDA [39], CSA [75], DATER [74, 77], and ranking SVM [9, 51].

We show the accuracies of the verification mode by the receiver operating characteristic (ROC) curve with z-normalization, which indicates the tradeoff between FAR and FAR for various genuine match acceptance threshold values. In addition, we show the accuracies of the identification mode by cumulative matching characteristic (CMC) curves in Fig. 6. In addition, as a summary of each mode's evaluation, we reported the EER with z-normalization and rank-1 identification rate in Table 2.

The results show that the proposed method yielded the best or the second best results for the both OUTD-B and OU-LP-Bag β data sets, which indicates the effectiveness of the proposed joint intensity and spatial metric learning for gait recognition under clothing and carrying status variations. Although Gabor GEI w/ RSM [15] achieved a superior rank-1 identification rate for OUTD-B, note that RSM cannot be applied in the verification mode because of its majority voting scheme for the gallery. It also cannot be used on a data set that includes a single sample per gallery (i.e., the OU-LP-Bag β data set). This is because it needs a regular within-class matrix for the gallery set. Therefore, considering the balance of application range and accuracy, we conclude that the proposed method is a suitable choice from among the benchmark methods.

5. Conclusion

This paper described a method of joint intensity metric learning and its application to robust gait recognition. We formulated the dissimilarity using a bilinear form of joint intensity and spatial metrics, and alternately optimize it by linear SVM or ranking SVM so as to well separate genuine and imposter pairs. We conducted experiments using OUTD-B and OU-LP-Bag β, and showed the effectiveness of the proposed method compared with other state-of-the-art methods.

Although the proposed method handles a common joint intensity metric over spatial positions, suitable joint intensity metrics may differ among the body parts (e.g., more leverage on the motion difference in the legs). Therefore, a future avenue of research is to extend the proposed framework to spatially dependent joint intensity metric learning. Moreover, because applications of the proposed framework are not limited to gait recognition, we plan to test the proposed method with other problems such as person re-identification.

Acknowledgement. This work was supported by JSPS Grants-in-Aid for Scientific Research (A) JP15H01693, the JST CREST "Behavior Understanding based on Intention-Gait Model" project, and Nanjing University of Science and Technology.

References

[1] 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.

[2] 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.

[3] R. Auckenthaler, M. Carey, and H. Lloyd-Thomas. Score normalization for text-independant speaker verification systems. *Digital Signal Processing*, 10(1-3):42–54, 2000.

[4] R. Baeza-Yates and G. Navarro. *A faster algorithm for approximate string matching*, pages 1–23. Springer Berlin Heidelberg, Berlin, Heidelberg, 1996.

[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.

[6] K. Bashir, T. Xiang, and S. Gong. Gait recognition without subject cooperation. *Pattern Recognition Letters*, 31(13):2052–2060, Oct. 2010.

[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.

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

[9] O. Chapelle and S. Keerthi. Efficient algorithms for ranking with svms. *Information Retrieval*, 13(3):201–215, 2010.

[10] 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.

[11] 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 76670Q–76670Q–13, 2010.

[12] M. Dundar, J. Theiler, and S. Perkins. Incorporating spatial contiguity into the design of a support vector machine classifier. In *2006 IEEE International Symposium on Geoscience and Remote Sensing*, pages 364–367, July 2006.

[13] 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.

[14] 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.

[15] 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.

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

[17] R. M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3(6):610–621, Nov 1973.

[18] M. Hashimoto, T. Fujiwara, H. Koshimizu, H. Okuda, and K. Sumi. Extraction of unique pixels based on co-occurrence probability for high-speed template matching. In *Optomechatronic Technologies (ISOT), 2010 International Symposium on*, pages 1–6, Oct 2010.

[19] M. Hofmann, J. Geiger, S. Bachmann, B. Schuller, and G. Rigoll. The tum gait from audio, image and depth (gaid) database: Multimodal recognition of subjects and traits. *J. Vis. Comun. Image Represent.*, 25(1):195–206, Jan. 2014.

[20] 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.

[21] 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.

[22] 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.

[23] Y. Kita. Change detection using joint intensity histogram. In *18th International Conference on Pattern Recognition (ICPR'06)*, volume 2, pages 351–356, 2006.

[24] Y. Kita. A study of change detection from satellite images using joint intensity histogram. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1–4, Dec 2008.

[25] Y. Kita. Background modeling by combining joint intensity histogram with time-sequential data. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 991–994, Aug 2010.

[26] T. Kozakaya, S. Ito, and S. Kubota. Random ensemble metrics for object recognition. In *The 13th IEEE International Conference on Computer Vision*, pages 1959 –1966, nov. 2011.

[27] W. Kusakunniran. Attribute-based learning for gait recognition using spatio-temporal interest points. *Image Vision Comput.*, 32(12):1117–1126, Dec. 2014.

[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.

[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.

[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.

[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.

[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.

[33] N. Kwak. Principal component analysis by l_p-norm maximization. *IEEE Transactions on Cybernetics*, 44(5):594–609, May 2014.

[34] J.-E. Lee, R. Jin, and A. K. Jain. Rank-based distance metric learning: An application to image retrieval. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8, June 2008.

[35] L. Lee. *Gait Analysis for Classification*. PhD thesis, Massachusetts Institute of Technology, 2002.

[36] 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.

[37] M. E. Leventon and W. E. L. Grimson. *Multi-modal volume registration using joint intensity distributions*, pages 1057–1066. Springer Berlin Heidelberg, Berlin, Heidelberg, 1998.

[38] X. Li, S. Maybank, S. Yan, D. Tao, and D. Xu. Gait components and their application to gender recognition. *Trans. on Systems, Man, and Cybernetics, Part C*, 38(2):145–155, Mar. 2008.

[39] K. Liu, Y. Cheng, and J. Yang. Algebraic feature extraction. *IEEE Trans. Circuits Syst. Video Technol*, 26(6):903–911, 2006.

[40] 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.

[41] 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.

[42] S. Lombardi, K. Nishino, Y. Makihara, and Y. Yagi. Two-point gait: Decoupling gait from body shape. In *Proc. of the 14th IEEE International Conference on Computer Vision (ICCV 2013)*, pages 1041–1048, Sydney, Australia, Dec. 2013.

[43] J. Lu and Y.-P. Tan. Uncorrelated discriminant simplex analysis for view-invariant gait signal computing. *Pattern Recognition Letters*, 31(5):382–393, 2010.

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

[45] Y. Makihara, T. Kimura, F. Okura, I. Mitsugami, M. Niwa, C. Aoki, A. Suzuki, D. Muramatsu, and Y. Yagi. Gait collector: An automatic gait data collection system in conjunction with an experience-based long-run exhibition. In *Proc. of the 8th IAPR Int. Conf. on Biometrics (ICB 2016)*, number O17, pages 1–8, Halmstad, Sweden, Jun. 2016.

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

[47] 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.

[48] 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.

[49] 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.

[50] R. Martín-Félez and T. Xiang. Gait recognition by ranking. In *Proceedings of the 12th European conference on Computer Vision - Volume Part I*, ECCV'12, pages 328–341, Berlin, Heidelberg, 2012. Springer-Verlag.

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

[52] 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.

[53] D. W. Mount. *Bioinformatics: Sequence and Genome Analysis*. Cold Spring Harbor Laboratory Press., 2nd edition, 2004.

[54] H. Murase and R. Sakai. Moving object recognition in eigenspace representation: Gait analysis and lip reading. *Pattern Recognition Letters*, 17:155–162, 1996.

[55] M. S. Nixon, T. N. Tan, and R. Chellappa. *Human Identification Based on Gait*. Int. Series on Biometrics. Springer-Verlag, Dec. 2005.

[56] J. H. Oh and N. Kwak. Generalization of linear discriminant analysis using l_p-norm. *Pattern Recognition Letters*, 34(6):679 – 685, 2013.

[57] N. Otsu. Optimal linear and nonlinear solutions for least-square discriminant feature extraction. In *Proc. of the 6th Int. Conf. on Pattern Recognition*, pages 557–560, 1982.

[58] M. Rokanujjaman, M. Islam, M. Hossain, M. Islam, Y. Makihara, and Y. Yagi. Effective part-based gait identification using frequency-domain gait entrophy features. *Multimedia Tools and Applications*, 74(9):3099–3120, May 2015.

[59] M. Rokanujjaman, M. Islam, M. Hossain, M. Islam, Y. Makihara, and Y. Yagi. Effective part-based gait identification using frequency-domain gait entrophy features. *Multimedia Tools and Applications*, 74(9):3099–3120, May 2015.

[60] 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.

[61] 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.

[62] 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.

[63] J. Shutler, M. Grant, M. Nixon, and J. Carter. On a large sequence-based human gait database. In *Proc. of the 4th Int. Conf. on Recent Advances in Soft Computing*, pages 66–71, Nottingham, UK, Dec. 2002.

[64] 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.

[65] 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.

[66] 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.

[67] 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.

[68] 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.

[69] 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.

[70] 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.

[71] 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.

[72] K. Q. Weinberger and L. K. Saul. Distance metric learning for large margin nearest neighbor classification. *J. Mach. Learn. Res.*, 10:207–244, June 2009.

[73] 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.

[74] 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.

[75] D. Xu, S. Yan, L. Zhang, H.-J. Z. andZhengkai Liu, and H.-Y. Shum. Concurrent subspaces analysis. In *Proc. of the IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, pages 203–208, Jun. 2005.

[76] 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.

[77] 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.

[78] S. Yu, D. Tan, and T. Tan. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In *Proc. of the 18th Int. Conf. on Pattern Recognition*, volume 4, pages 441–444, Hong Kong, China, Aug. 2006.

[79] S. Yu, T. Tan, K. Huang, K. Jia, and X. Wu. A study on gait-based gender classification. *IEEE Trans. on Image Processing*, 18(8):1905–1910, Aug. 2009.

[80] 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 (FGR06)*, pages 529–534, April 2006.

[81] W. S. Zheng, S. Gong, and T. Xiang. Person re-identification by probabilistic relative distance comparison. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 649–656, June 2011.