



Speed-Invariant Gait Recognition Using Single-Support Gait Energy Image

Chi Xu^{1,2}  · Yasushi Makihara² · Xiang Li^{1,2} · Yasushi Yagi² · Jianfeng Lu¹

Received: 30 June 2018 / Revised: 20 February 2019 / Accepted: 29 April 2019 /
Published online: 11 June 2019
© The Author(s) 2019

Abstract

Gait is one of the most popular behavioral biometrics because it can be authenticated at a distance from a camera without subject cooperation. Speed differences between matching pairs, however, cause significant performance drops in gait recognition, and gait mode difference (i.e., walking versus running) makes gait recognition further challenging. We therefore propose a speed-invariant gait representation called single-support GEI (SSGEI), which realizes a good trade-off between speed invariance and stability by aggregating multiple frames around single-support phases. In addition, to mitigate the pose differences between walking and running modes at single-support phases, we morph walking and running SSGEIs into intermediate SSGEIs between walking and running mode, where we exploit a free-form deformation field from the walking or running modes to the intermediate mode obtained by training data. We finally apply Gabor filtering and spatial metric learning as postprocessing for further accuracy improvement. Experiments on two publicly available datasets, the OU-ISIR Treadmill Dataset A and the CASIA-C Dataset demonstrate that the proposed method yields the state-of-the-art accuracies in both identification and verification scenarios with a low computational cost.

Keywords Gait recognition · Single-support phase · Speed invariance · Morphing

✉ Chi Xu
xuchisherry@gmail.com

Yasushi Makihara
makihara@am.sanken.osaka-u.ac.jp

Xiang Li
lixiangmzx@gmail.com

Yasushi Yagi
yagi@am.sanken.osaka-u.ac.jp

Jianfeng Lu
lujf@njust.edu.cn

¹ School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

² Department of Intelligent Media, The Institute of Scientific and Industrial Research, Osaka University, Osaka 567-0047, Japan

1 Introduction

Gait recognition is a technique to authenticate a person from his/her walking style and has advantages over other physiological biometric modalities (e.g., DNA, fingerprints, irises, and faces) in terms that it works even with relatively low-resolution images [32] (e.g., CCTV footage captured at a large distance without subject cooperation), and that gait is difficult to obscure and imitate. The demand for gait recognition has therefore grown in many applications for surveillance and forensics [1, 9, 24].

However, involving uncooperative subjects makes gait recognition more challenging, as gait may be influenced by various covariates such as views, shoes, surfaces, clothing, carriage, and speed [2, 34]. Among these covariates, speed change is one of the most common challenging factors and often occurs in real scenes depending on the situation (e.g., a perpetrator running from a crime scene). Gait recognition performance may significantly degrade under speed variation because the speed change induces changes in appearance-based gait features (e.g., the gait energy image (GEI) [6] and frequency-domain features [26]), which are often used in the gait recognition community. Extensive efforts to achieve speed-invariant gait recognition have therefore been made, as in the previous studies [5, 16, 17, 23, 27]. While these approaches can mitigate the effect of speed on gait recognition to some extent, most of them work poorly under large speed changes, or suffer from high computational cost, which is an important aspect in real-world applications.

In contrast to the above work, there are approaches to cross-speed gait recognition using dynamic part attenuation because the speed change mainly affects dynamic parts like arm swing and stride length.

Tanawongsuwan and Bobick [41] proposed using silhouettes at the single-support phases as a part of the gait features because the single-support phases, where the limbs are the most closed, as shown in Fig. 1, are not drastically varied as the speed changes, while double-support phases significantly change depending on speed. However, a single key-frame at single-support phase is easily influenced by silhouette segmentation noise, temporary posture changes, and phase estimation errors, which also drop the gait recognition performance.

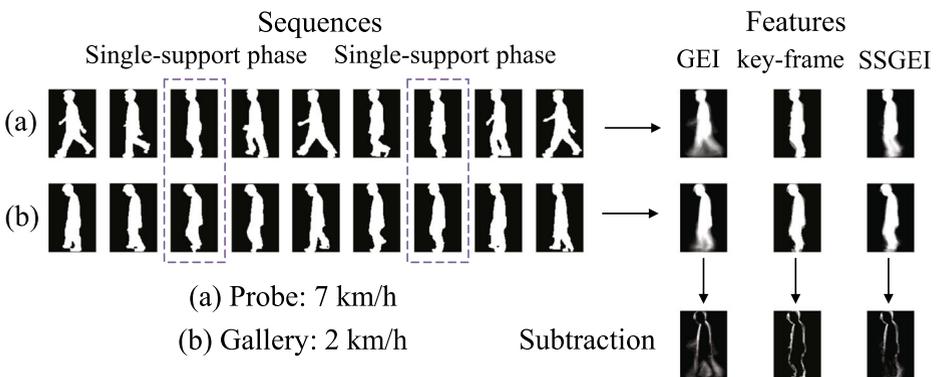


Fig. 1 Examples for comparing GEIs, key-frames, and SSGEIs. Nine frames are evenly chosen from a period in both the gallery (2 km/h) and probe (7 km/h) sequences of the same subject. The corresponding GEIs, single key-frame at the single-support phase, and SSGEIs are shown on the right. The subtraction image for each feature is shown at the bottom. It is clear that the SSGEIs can reduce the appearance differences caused by speed variance, posture change, and phase differences simultaneously

Iwashita et al. [12] applied a mutual subspace method (MSM) to cross-speed gait recognition, in which a set of silhouettes at various phases is represented by a subspace and a dissimilarity measure is obtained as the canonical angle between the subspaces for a matching pair. More specifically, in [12], a silhouette at an arbitrary phase, i.e., a transition from a single-support phase to double-support phase, is well approximated on each subspace and the canonical angle is obtained by minimizing the forming angle between a pair of silhouettes represented in each subspace. Therefore, in the cross-speed scenario, the canonical angle is expected to be obtained at the single-support phases, where the effect of the speed changes is minimal. However, the possibility of a false match at phases other than the single-support phase remains because the subspace represents silhouettes at various phases.

To overcome these defects, we propose a speed-invariant and stable gait representation called single-support GEI (SSGEI) for speed-invariant gait recognition. By combining single-support phases with the concept of GEI [6], in which multiple frames are aggregated for silhouette noise reduction, we also aggregate multiple frames of a certain duration around the single-support phase. Because longer duration leads to more stability but less speed invariance, while shorter duration leads to less stability but more speed invariance, we determine the optimal duration that balances the speed invariance and stability using a training set.

In addition to the above cross-speed gait recognition within a walking mode, there are a few studies [5, 8, 48] on those within a running mode or between walking and running modes (i.e., cross-mode gait recognition). In particular, cross-mode gait recognition is much more challenging than recognition within the same mode, because even the proposed SSGEIs are subject to changes in body inclination angle and leg motion between walking and running modes (see the running probe SSGEI (8 km/h) and walking gallery SSGEI (7 km/h) in Fig. 2).

Of these studies, only Guan and Li's work [5] tackled cross-mode gait recognition to the best of our knowledge. The method in [5], however, works poorly for cross-mode gait recognition because they apply a common metric learning technique called the random subspace method (RSM), regardless of the mode (i.e., walking or running) and hence they cannot absorb the large differences between the walking and running modes.

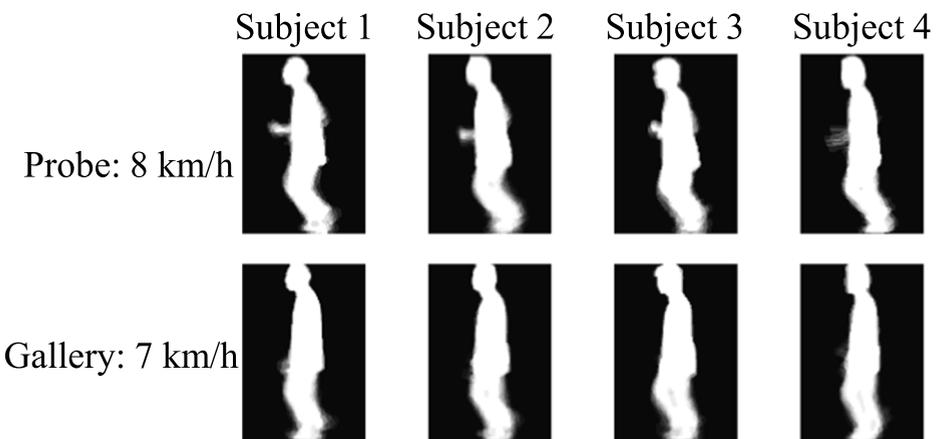


Fig. 2 SSGEIs of walking and running sequences from different subjects. Each column shows the SSGEIs of a running probe sequence (8 km/h) and walking gallery sequence (7 km/h) from the same subject

Taking a closer look at the SSGEIs between the walking and running shown in Fig. 2, we note that there are some common changes in the upper bodies and leg motions among different subjects, and hence it is reasonable to consider a geometric transformation to register the pose differences between different gait modes, which is a widely used preprocessing procedure in face recognition community, and has not been introduced in the field of cross-speed gait recognition to the best of our knowledge. We therefore generate a generic warping between the walking and running modes across the population to cope with the cross-mode gait recognition. The contributions of this work are four-fold¹.

1. **A speed invariant and stable gait representation.** The proposed SSGEI realizes a good trade-off between speed invariance and stability, which can be intuitively understood by the example shown in Fig. 1. The subject exhibits temporary posture changes as he looks down in several frames in the gallery sequence (2 km/h), while he keeps on walking normally in the probe sequence (7 km/h). Moreover, the selected single-support key-frames may contain slight phase differences. In contrast, GEIs mitigate the temporary posture changes and phase differences, but are directly affected by the changes in dynamic parts like stride and arm swing due to speed variation. Consequently, both key-frames and GEIs suffer from significant differences between gallery and probe ones, which may lead to false matches. However, these differences between gallery and probe images are well suppressed with the proposed SSGEI owing to its balance between speed invariance and stability, which are derived from the concepts of key-frames at the single-support phase and aggregation in the GEI, respectively.
2. **General framework for three speed-invariant gait recognition cases.** We designed a general framework based on the SSGEI to appropriately handle three different scenarios of speed-invariant gait recognition, i.e., within-walking, within-running, and cross-mode scenarios. More specifically, we first define two gait mode classes (i.e., walking and running) considering the trade-off between fine warping fields and difficulties in the gait mode classification, and then apply a mode classification technique and subsequently compensate for the cross-mode difference by morphing the SSGEIs of walking and running modes into those of the intermediate mode by free-form deformation (FFD) [35], which is brought into speed-invariant gait recognition for the first time.
3. **State-of-the-art accuracy for speed-invariant gait recognition on two publicly available data sets.** We evaluated the proposed method in within-walking, within-running, and cross-mode scenarios with the OU-ISIR Gait Database, Treadmill Data set A [25]. We also tested the within-walking scenario with the CASIA Gait Database, Data set C [38]. The former dataset contains the largest speed variations, and is the only data set that includes running sequences, while the latter dataset contains a larger number of subjects with speed variations in the walking mode, which makes the performance evaluation more statistically reliable. The experimental results for both datasets show that the proposed method yields the state-of-the-art accuracies both in terms of verification and identification scenarios.

¹This paper is an extended version of the conference paper [46]. More specifically, the second contribution (methodological extension) and parts of the third and fourth contributions (experimental extension with respect to data sets and scenarios such as within-running and cross-mode matching) are the extensions.

4. **Low computational cost.** The proposed method is also executable with a low computational cost and hence is more suitable for real-world surveillance applications, while the other state-of-the-art methods require relatively high computational costs.

2 Related work

2.1 Speed-invariant gait recognition in the within-walking scenario

Currently, various speed-invariant gait recognition methods have been proposed, and they fall into two categories [16]: i) transforming features from a reference speed to another speed and ii) extracting speed-invariant gait features. The core of the first category is to learn the relationship between features under different walking speeds [17], such as stride normalization for double-support frames [41] and factorization-based speed transformation model [27]. However, the transformation-based approaches suffer from high-computational model fitting and perform relatively poorly when the speed change is large.

In the second category, speed-invariant gait features are employed for gait recognition [11, 12, 14, 16, 23, 39, 40]. For example, in [16], based on Procrustes shape analysis descriptors, the differential composition model (DCM) was introduced to differentiate the effects on each body part caused by speed change. Iwashita et al. [12] applied a mutual subspace method (MSM) to a set of gait silhouette images and a canonical angle between the gallery and probe subspaces is computed as dissimilarity measure, which is often chosen from the single-support phases in the cross-speed case. They further extended this approach in [11] by dividing the human body into multiple areas, and using a matching weight to select the relatively static parts. The elimination of dynamic parts helps to reduce the effects of walking speed variations, but it may fail if temporary posture changes occur on the static parts. On the other hand, it is unsuitable to be extended to the cross-mode gait-recognition, where both static and dynamic parts obviously change across the walking and running modes.

Another direction is to directly apply a metric learning approach to cross-speed gait recognition. Guan and Li [5] employed RSM to combine a large number of weak classifiers, which can reduce the generalization errors caused by different walking speeds. The RSM framework achieves significant performance improvements, but it faces two limitations: 1) the accuracy varies because of its random nature and 2) it is time-consuming to calculate because it needs to construct a large number of random subspaces and execute a matching process for each one.

2.2 Cross-mode gait recognition

Most cross-speed gait recognition studies only focus on within-walking cases, yet running gait recognition, particularly cross-mode gait recognition, is worth further investigation because a running perpetrator may often need to be recognized only with his/her walking gallery in real scenes. Yam et al. [48] proposed an analytical model using the biomechanics of human locomotion and a unique mapping was found between walking and running gait features for each subject. A generic mapping across the population may, however, not exist, which limits its use in surveillance for identifying unknown runners by their walking features only. In a unique study that evaluates both the speed changes in each mode and the

cross-mode scenario, Guan and Li [5] applied RSM. Although they achieved high accuracies in within-walking and within-running scenarios, the method still performed poorly in cross-mode tasks.

2.3 Gait mode classification

There has been considerable interest in the classification of gait modes or, more generally, of different types of human actions [4]. Yu et al. [50] used a three-layer feedforward network to classify walking, running, and other action types based on the trajectories in eigenspace. Cheng et al. [3] computed a characteristic frequency using the mean motion magnitude between frames. Kim et al. [13] proposed a tensor canonical correlation analysis method. In [21], a real-time human action recognition solution was proposed based on luminance field trajectory analysis and learning. Fihl et al. [4] introduced the duty-factor (i.e., the fraction of the stride duration over which each foot remains on the ground) to characterize gait modes, which is independent of challenging factors such as varying speeds. Although some of the existing methods have a high classification accuracy, they still require relatively high computational costs.

2.4 Deep learning-based gait recognition

To date, the deep learning-based approaches have demonstrated the state-of-the-art performance in gait recognition, which mainly focus on tackling gait recognition under view variations. While [44] and [43] proposed deep convolutional neural network (CNN) models using raw silhouette images as the inputs, Shiraga et al. [36] designed GEINet whose input is a single GEI. Some latest works [45, 51] presented the CNN models with two inputs, where the similarities of these two inputs were learnt to discriminate between the same subject pairs and different subject pairs, and in [37], CNN architectures with different input and output were explored for gait verification and identification scenarios respectively. These approaches achieved superior performance in comparison to traditional methods, sufficiently enormous number of training samples, however, are required to obtain reliable CNN models, which are unsuitable to be applied for datasets with small sample size.

3 Gait recognition using SSGEI

3.1 Overview

An overview of the proposed framework is shown in Fig. 3. Given a matching pair of gait silhouette sequences, i.e., a gallery and probe, which can be extracted from raw images by a background subtraction-based graph-cut segmentation [28], or recent state-of-the-art deep learning-based semantic segmentation methods such as RefineNet [22], we first generate the size-normalized and registered silhouette sequences by height normalization and registration using the region center [26]. After detecting the gait period, by aggregating multiple single-support frames over the optimal duration of the period, we extract the SSGEI as a gait feature.

Because a subsequent procedure is changed depending on the gait mode (walking or running) for the matching pair (and because the gait mode is not known in advance), we estimate the gait modes based on SSGEIs using a gait mode classifier. In the cross-mode case (i.e., one SSGEI is walking and the other is running), to reduce appearance changes

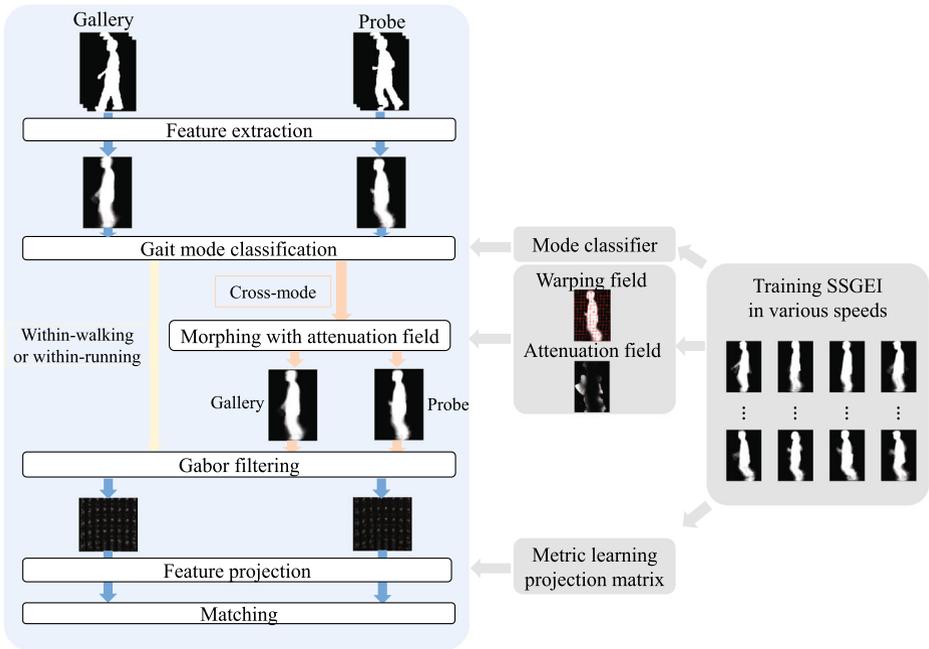


Fig. 3 Overview of the proposed framework

caused by the pose difference between walking and running, we morph both the gallery and probe SSGEIs into intermediate poses using a generic warping field. In addition, because residuals may still remain after the generic warping because of the subject-dependent transition between the gait modes (e.g., some subjects may raise their arms higher in running mode than generic subjects, even if their arm swings are similar in walking mode), we attenuate the SSGEIs at such easily affected positions. We finally apply Gabor filtering and metric learning to the obtained SSGEIs as postprocesses and then compute the L2 distance as the dissimilarity measure. The final performance for verification scenarios (i.e., one-to-one matching) is obtained by comparing the dissimilarity score with an acceptance threshold, while the accuracy for identification scenarios (i.e., one-to-many matching) is calculated using nearest neighbor classifier, which is the most widely used classifier in gait identification community.

Details for the procedures are given in the following sections.

3.2 SSGEI extraction

3.2.1 Representation

A gait period is first detected from the lower body parts of the normalized silhouette sequence. Given the body height H , we set the vertical position of the knee to $0.285H^2$ based on anatomical data statistics, as suggested in [7]. A temporal series of the width of the

²In this coordinate system, the vertical positions of the foot bottom and the head top are denoted as 0 and H , respectively.

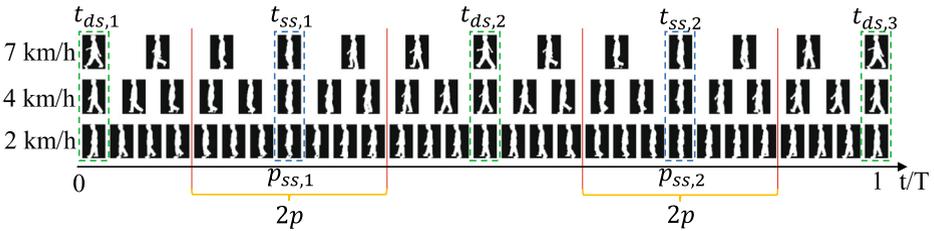


Fig. 4 Demonstration of the gait period and duration for an SSGEI along with the normalized silhouette sequences from three different walking speeds. The horizontal axis t/T represents a non-dimensional time normalized by gait period T . Note that the frame intervals for three walking speeds are different because of the gait period difference. Two single-support phases are denoted as $p_{ss,k}$ ($k = 1, 2$) in this non-dimensional time domain. The durations composed of multiple single-support phases within the range $[p_{ss,k} - p, p_{ss,k} + p]$ ($k = 1, 2$) are selected for constructing the SSGEI, where p is a hyperparameter for duration selection

lower body from the foot bottom to the knee is computed and the local maxima and minima are found as the double-support phases and single-support phases, respectively.

Thus, we define a gait period of T [frames] to start from a double-support phase ($t = t_{ds,1} = 0$), then go through two single-support phases ($t = t_{ss,1}$ and $t = t_{ss,2}$) as well as another in between double-support phase ($t = t_{ds,2}$), and finally end with the third double-support phase ($t = t_{ds,3} = T$), which is shown in Fig. 4.

To define the duration around single-support phases in a walking speed rate-invariant way, we convert a time $t \in \mathbb{Z}$ [frames] into a non-dimensional time $p = t/T \in \mathbb{R}$, which is normalized by period T . Suppose that we take a $2p$ duration around each single-support phases $p_{ss,k}$ ($k = 1, 2$) in the non-dimensional time domain. Then, the duration around the k -th single-support phase is defined as $[p_{ss,k} - p, p_{ss,k} + p]$. Note that the duration parameter p is subject to $0 < p \leq 1/4$ (the duration will cover the whole period if $p = 1/4$).

Once the durations are defined, we can convert them back into the original time domain and obtain the starting and ending frames for the k -th duration as $t_{ss,k}^s(p) = \lceil (p_{ss,k} - p)T \rceil$ and $t_{ss,k}^e(p) = \lfloor (p_{ss,k} + p)T \rfloor$, respectively, where $\lceil \cdot \rceil$ and $\lfloor \cdot \rfloor$ are ceiling and floor functions, respectively.

An SSGEI can now be computed based on the durations. Let a binary silhouette value at position (x, y) from the t -th frame in the size-normalized silhouette sequence be $I(x, y, t)$, where 0 and 1 indicate the background and foreground, respectively. SSGEI $S(x, y; p)$ is defined using duration parameter p as

$$S(x, y; p) = \frac{1}{2} \sum_{k=1}^2 \frac{1}{t_{ss,k}^e(p) - t_{ss,k}^s(p) + 1} \sum_{t=t_{ss,k}^s(p)}^{t_{ss,k}^e(p)} I(x, y, t). \tag{1}$$

Examples of SSGEIs can be found in Fig. 1. The SSGEI shows its effectiveness clearly when compared with GEI and a single single-support key-frame.

3.2.2 Optimal duration estimation

We next need to carefully select optimal duration parameter p to realize a good trade-off between the speed invariance and stability for the proposed SSGEI. Consequently, we introduce a well-known criterion for discrimination capability, i.e., the Fisher ratio of between-class distance and within-class distance using a training set including speed variations. Note that composition of the training set varies depending on the scenario such as

within-walking, within-running, and cross-mode matching. More specifically, for within-walking or within-running scenarios, the training set only includes walking or running SSGEIs, respectively, while for the cross-mode case, the training set includes morphing results at the intermediate pose of walking and running SSGEIs, which is introduced in later sections. As a result, the optimal duration parameter p^* is obtained to maximize the Fisher ratio of the between-class distances and within-class distances. We refer readers to [46] for more details about the acquisition of the Fisher ratio.

3.3 Classification of walking and running

Firstly, two gait modes, i.e., walking and running, are defined based on the walking/running pose of a subject, which mainly differs in body inclination angle and leg motion (see Fig. 2). Because dynamic part variation caused by speed changes within the same gait mode may degrade the classification accuracy of walking and running modes, we adopt the SSGEI to address this problem by considering the trade-off between speed invariance and stability within the same mode. Note that we can see relatively common (i.e., subject-independent) pose changes between walking and running modes, as shown in Fig. 2. Moreover, because the gait period of the running mode is generally much shorter than that of the walking mode, we exploit the gait period T [frames] as a useful feature for gait mode classification. More specifically, we define a concatenated feature vector of the gait period T [frames] and SSGEI and feed it to a linear support vector machine for classification into walking or running mode. Note that the gait mode classifier is trained using the training set composed of walking and running SSGEIs with diverse speeds.

3.4 Morphing by FFD

To overcome the large intra-class differences between running and walking modes, we utilize a generic warping field between them across the population. For this purpose, we utilize the notion of FFD with piece-wise linear interpolation, because the FFD provides a high degree of flexibility for describing the transformation of non-rigid objects such as a human, as well as maintains derivative continuity at adjacent regions [35], i.e., gait characteristics for person authentication, unlike some other example-based view transformation approaches such as [15, 26, 30] may corrupt the geometric continuity of the gait features.

Instead of a conventional bi-directional cost function to minimize the error between the target and transformed source as well as between the source and inverse-transformed target [19], we introduce a cost function to minimize the error between targets and sources that are both transformed into intermediate SSGEIs. More specifically, we allocate a set of control points on the SSGEI and then define a set of two-dimensional displacement vectors from the walking to intermediate SSGEI on the control points as \vec{u} . We then define a warping field $F(\vec{u})$ from walking SSGEI to intermediate SSGEI by piece-wise linear interpolation. Similarly, we consider a reverse version of the displacement vector $-\vec{u}$, and its warping field $F(-\vec{u})$ from running SSGEI to intermediate SSGEI. We finally match the morphed SSGEIs in the intermediate domain. The advantages of this deformation representation are i) the deformation between walking and running is treated symmetrically and ii) the degree of deformation from each walking and running mode to the intermediate mode is equal to each other (i.e., $\|\vec{u}\| = \|-\vec{u}\|$).

Using the above concept, we denote a pair of source and target SSGEIs (i.e., running and walking SSGEIs) as $S_{i,j}^S, S_{i,j}^T \in \mathbb{R}^{H_S \times W_S}$ ($i = 1, \dots, N_c, j = 1, \dots, k_i$), respectively, where N_c and k_i are the number of training subjects and source/target pairs for the i -th

training subject, respectively. Suppose a mapping of the warping field from the source to the intermediate mode is obtained as $F(\vec{u})$ by a piece-wise linear interpolation of \vec{u} . Then, the transformed source SSGEI is represented as $S_{i,j}^S \circ F(\vec{u})$, where \circ indicates a transformation operator. Similarly, the transformed target SSGEI is represented as $S_{i,j}^T \circ F(-\vec{u})$. Consequently, we obtain the optimal displacement vector \vec{u}^* by minimizing the summation of differences between the transformed source and target GEIs in the intermediate domain as

$$\vec{u}^* = \arg \min_{\vec{u}} E(\vec{u}), \quad (2)$$

where

$$E(\vec{u}) = \sum_{i=1}^{N_c} \sum_{j=1}^{k_i} \|S_{i,j}^S \circ F(\vec{u}) - S_{i,j}^T \circ F(-\vec{u})\|_F^2 + \lambda R(\vec{u}). \quad (3)$$

Here, $R(\vec{u})$ is a smoothness term, i.e., a linear elastic constraint on the displacements between adjacent control points [19] and λ is a hyperparameter to control the smoothness. We solve the optimization of (2) by gradient descent. More specifically, the displacement vectors \vec{u} are set to be zero at initialization, and then the gradient descent of $E(\vec{u})$ is computed to update \vec{u} iteratively until convergence. As such, we obtain the intermediate SSGEIs $S_{i,j}^{S'} = S_{i,j}^S \circ F(\vec{u}^*)$ and $S_{i,j}^{T'} = S_{i,j}^T \circ F(-\vec{u}^*)$ transformed from the source and the target, as shown in Fig. 6a–f.

3.5 Attenuation field

While the above generic warping mitigates the intra-subject inter-mode differences to some extent, residuals may still remain because of subject-dependent transitions between the gait modes (e.g., some subjects may raise their arms higher in running mode than generic subjects even if their arm swings are similar in walking mode.) as shown in Fig. 2. We therefore introduce an attenuation field to suppress such subject-dependent residuals.

For this purpose, we further employ an outlier detection method [31] in a framework of transportation minimization-based morphing called the earth mover's morphing framework [29] to determine appearing/disappearing regions between a source and target derived from the subject-dependent residuals. We refer the reader to [29, 31] for more details. Specifically, given a transformed source and target SSGEIs $S_{i,j}^{S'}$ and $S_{i,j}^{T'}$, we regard the brightness at each pixel in the transformed SSGEIs as a sort of mass assigned to the pixel and then try to transport all the pixels in the transformed source SSGEI into those in the transformed target SSGEIs with the minimal cost (i.e., the weighted sum of travelling distances by mass). Here, the subject-dependent residual such as arm swing difference from the generic warping may require a large transportation cost, and hence we prepare an exceptional path to a trash bin, which is automatically assigned to a pixel whose transportation cost exceeds a certain threshold in the transportation minimization framework. Consequently, we regard such trash-bin pixels as outliers for the generic warping field between walking and running modes, and then construct an attenuation field by aggregating the trash-bin pixels over all the training subjects, as shown in Fig. 6i.

Once the attenuation field is obtained, we attenuate the intensities of both the transformed source and target SSGEIs for each pixel, e.g., if an attenuation value at a certain pixel is 70%, the intensities of the transformed SSGEIs at the same pixel are reduced by 70%.

3.6 Update morphing

To mitigate the effect of the subject-dependent residuals in the acquisition process of the generic warping field, we recompute the optimal displacement vector \vec{u}^* by introducing the attenuation field. More specifically, when computing the Frobenius norm in (3), we reduce the intensities of both the transformed source and target SSGEIs for each pixel depending on the attenuation field. As such, we obtain the updated displacement vector, and then update the attenuation field in turn.

4 Postprocessing

The effectiveness of Gabor filtering has been demonstrated in the context of biologically inspired image understanding processes [18, 42] and its effectiveness in gait recognition has been also demonstrated in [5, 42, 47]. We therefore also introduce Gabor filtering as a post-processing step for the proposed SSGEIs in the within-walking and within-running cases, as well as for the transformed SSGEIs in the cross-mode case (referred to as Gabor-SSGEI). In the Gabor feature space, we further employ two-dimensional principle component analysis (2DPCA) [49] to reduce the feature dimensions in the column direction while retaining 99% of the variance in our applications. Two-dimensional linear discriminant analysis (2DLDA) [20] is then exploited to obtain a discriminative projection in the row direction. We refer readers to the supplemental material for the details of postprocessing.

5 Experiments

5.1 Data sets and parameter settings

To evaluate the proposed method, we adopted two publicly available data sets, i.e., the OU-ISIR Gait Database, Treadmill Dataset A (OUTD-A) [25] and CASIA Gait Database, Dataset C (CASIA-C) [38].

The first data set contains image sequences of 34 subjects with speed variations ranging from 2 km/h to 10 km/h in 1 km/h intervals. We use this dataset to evaluate our method for all the following experiments because of its largest speed variations, and because it is the only data set that includes running sequences. Following the settings of the dataset [25], walking speeds from 2 km/h to 7 km/h are used for the within-walking case, while running speeds from 8 km/h to 10 km/h are used for the within-running case. The cross-mode case includes all speeds, where the galleries are walking speeds while the probes are running speeds and vice versa. Nine subjects were used for training parameter p , the generic warping field, and 2DPCA and 2DLDA. The other disjoint 25 subjects were used for testing according to the protocol suggested in [27]. In identification scenarios, we followed an uncooperative setting, i.e., subjects in a specific gallery may have different speeds and/or gait modes, which makes the identification task more challenging than the cooperative setting, i.e., subjects in a specific gallery have the same speed and gait mode. Therefore, gait mode classification was applied to both probe and gallery sequences before subsequent procedures.

The second data set is composed of 153 subjects with three different walking speeds, i.e., slow (fs), normal (fn), and fast (fq) walking. This data set was used for experiments in Section 5.7 to make the performance evaluation more statistically reliable. Following [17], 33 subjects were randomly selected to make up the training set, and the rest of the 120

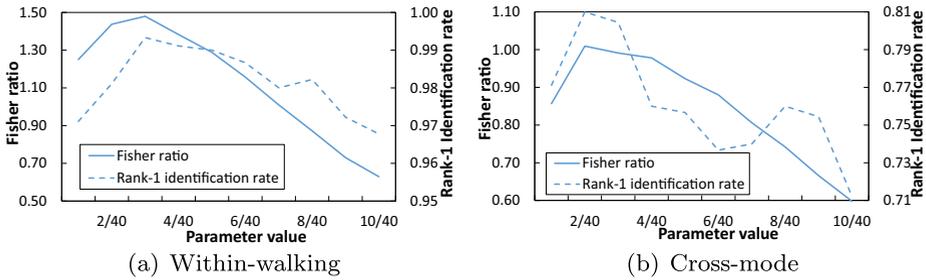


Fig. 5 Fisher ratio and corresponding rank-1 identification rate of the testing set for each duration parameter candidate using the training set

subjects were used for the testing set. To mitigate the effect of the random selection on performance evaluation, we repeated this random selection processes 10 times, and report the mean accuracies for the identification scenarios and accuracies in the verification scenarios using the entire set of dissimilarity scores. Eight sequences were collected for each subject, which were composed of four *fn* sequences, two *fs*, and two *fq* sequences. Three *fn* sequences, one *fs* sequence, and one *fq* sequence were chosen as the *fn*, *fs*, and *fq* galleries, respectively, while the other sequences were used as probes. For example, when three *fn* sequences were used as the *fn* gallery, the remaining one *fn*, two *fs*, and two *fq* sequences were probes.

In our applications, the dimensions of 2DLDA were all chosen within the range $[1, 10, 20, \dots, 250]$ to maximize the accuracy of the training set in both the verification and identification scenarios.

5.2 Analysis on the optimal duration parameter

As described in Section 3.2.1, the optimal duration parameter p is selected within $0 < p \leq 1/4$. Concretely speaking, we empirically prepared a discrete set of parameter candidates as $p \in \{i/40\} (i = 1, 2, \dots, 10)$ at $1/40$ intervals (when $p = 10/40$, the duration includes the whole period). We report the Fisher ratio of the training set as well as the corresponding rank-1 identification rate (i.e., identification accuracy) of the testing set for each parameter candidate p under within-walking and cross-mode cases in Fig. 5. We refer readers to the supplemental material for the within-running case. Note that for the cross-mode case, parameter p was chosen using the transformed SSGEIs. These results show that the best rank-1 identification rates are obtained at the optimal durations by the Fisher ratios for all the three cases, which shows the generality of duration parameter p . As a result, we adopted $p^* = 3/40, 8/40$, and $2/40$ in our experiments for the within-walking, within-running, and cross-mode cases, respectively.

5.3 Gait mode classification

Because it is an important preprocessing step, we report gait mode classification accuracy. For comparison, we also tested GEI and the key-frame at the single-support phase (simply called the key-frame) concatenated with the gait period in addition to the proposed SSGEI. A training set for the gait mode classification was composed of sequences with multiple periods under nine speeds from the nine training subjects, which summed up to 279 samples, while a testing set from the other 25 test subjects contained 775 samples.

Table 1 Correct classification rate [%] of the gait mode using GEI, key-frame, and SSGEI for each subset

Feature \ Subset	GEI	key-frame	SSGEI
Training	100	100	100
Gallery in test	<i>99.87</i>	96.77	100
Probe in test	<i>98.97</i>	96.00	100

Bold and italic bold indicate the best and second-best accuracies. This font convention is used to indicate performances throughout this paper

The results in Table 1 show that only SSGEI yields 100% correct classification rates for all three subsets, and hence we can avoid the accuracy decrease due to the misclassification of the gait modes. In other words, the high correct classification rate implies the existence of a common transformation between walking and running SSGEIs among different subjects, and hence indicates the technical soundness of using a generic warping field between walking and running SSGEIs across the population.

5.4 Visualization of morphing process

To better understand the effectiveness of the transformed SSGEI, we visualize the morphing process using two typical examples, i.e., the easiest cross-mode case of running at 8 km/h and walking at 7 km/h and the most difficult cross-mode case of running at 10 km/h and walking at 2 km/h, as shown in Fig. 6.

Given a pair of running and walking SSGEIs (Fig. 6a and d), they are transformed into intermediate SSGEIs (Fig. 6c and f) using the generic warping fields from running (Fig. 6b) and from walking (Fig. 6e). Because of pose differences between the original running and walking SSGEIs, there are relatively large residuals (Fig. 6g) in both examples. By transforming them with the generic warping fields (Fig. 6b and d) as well as the attenuation field (Fig. 6i), the residuals are significantly reduced (Fig. 6h), which illustrates the effectiveness of the morphing techniques.

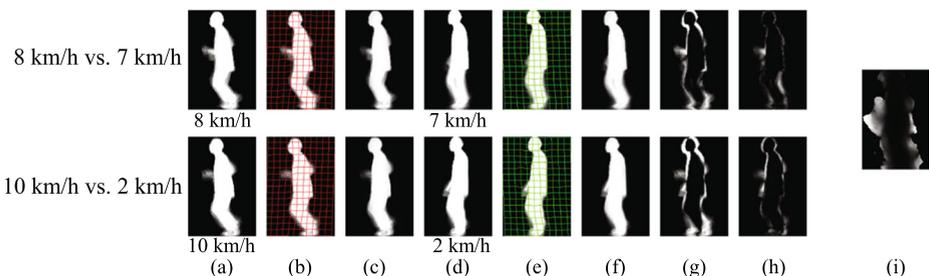


Fig. 6 Two examples of transformed SSGEIs. **a** Original running SSGEI. **b** Generic warping field from running to intermediate SSGEI. **c** SSGEI transformed from (a). **d** Original walking SSGEI. **e** Generic warping field from walking to intermediate SSGEI. **f** SSGEI transformed from (d). **g** Subtraction of original SSGEIs (a) and (d). **h** Subtraction of transformed SSGEIs (c) and (f). **i** Visualization of the attenuation field. In (i), a brighter value indicates more attenuation. Most of the arm regions as well as partial front and back leg regions are highly attenuated, which is consistent with our intuition that the transition of arm swings and leg motions between walking and running modes is highly dependent on the subject

Table 2 Overall rank-1 identification rate (denoted as Rank-1)[%], EER with and without z-normalization (denoted as z-EER and EER)[%] of key-frame, GEI, and SSGEI over all combinations of speeds in the probe and gallery for all three cases

(a) Within-walking and within-running cases						
	Within-walking			Within-running		
	key-frame	GEI	SSGEI	key-frame	GEI	SSGEI
Rank-1	73.8	62.6	80.3	83.1	91.6	94.2
z-EER	12.1	15.0	9.0	8.4	5.1	4.0
EER	16.7	27.9	13.6	11.1	8.0	5.8

(b) Cross-mode case						
	key-frame	GEI	SSGEI	Morphed key-frame	Morphed GEI	Morphed SSGEI
Rank-1	7.1	8.8	8.4	37.3	40.9	43.9
z-EER	45.0	38.4	40.9	23.9	20.2	19.8
EER	45.6	41.6	43.2	27.1	26.9	25.0

5.5 Feature comparison

In this section, three features, key-frame, GEI, and SSGEI, were tested with OUTD-A before the postprocessing steps of Gabor filtering and metric learning were applied for all three cases, within-walking, within-running, and cross-mode cases. We evaluated the accuracies in identification and verification scenarios using the rank-1 identification rate and equal error rate (EER) of the false acceptance rate (FAR) and false rejection rate (FRR), respectively.

First, the accuracies for both verification (EER with and without z-normalization [33]) and identification scenarios in the within-walking case are shown in the left columns of Table 2(a). Because the within-walking case contains various speed changes between the probe and gallery, GEI performs the worst as it is very sensitive to the walking speed change. Key-frame yields the second-best accuracy, as this method uses frames at the single-support phases, which are insensitive to speed changes, but it is less stable at the same time. In contrast, the proposed SSGEI feature achieves the best accuracy for both the verification and identification scenarios.

Similarly, we show the accuracies in the within-running case in the right columns of Table 2(a). Because the running speed variation in OUTD-A is smaller (i.e., from 8 km/h to 10 km/h) than the within-walking case (2 km/h to 7 km/h) and most of subjects increase their speed by shortening their gait periods rather than widening their stride length, appearance variation caused by speed changes is limited. The stability is therefore more important than the speed invariance in the within-running case. Hence, the key-frame method performs worse than GEI in the within-running case because it sacrifices the stability by aggregating only two single-support frames. Finally, the proposed SSGEI still yields the highest accuracy among the three features.

As for the cross-mode case, to demonstrate the effectiveness of the morphing process, we compared the above mentioned three features without and with morphing. As shown in Table 2(b), all three features perform poorly without morphing because of the large pose differences between walking and running. After the morphing procedure, the accuracies of the three features significantly improve, and the morphed SSGEI achieves the best performance.



Fig. 7 Matching examples of six features in the cross-mode case. The gallery and probe speeds are 7 km/h and 8 km/h, respectively. (a) Probe. (b) False match in the gallery (imposter). (c) True match in the gallery (genuine). (d) Subtraction image for the false match. (e) Subtraction image for the true match. The red bounding box indicates the subtraction with a smaller Euclidean distance

For a more intuitive understanding, we present typical examples of the above six features in Fig. 7 with a pair of true and false matches and their corresponding subtraction images. The subtraction images illustrate that the morphing procedure greatly reduces the differences between walking and running features. However, morphing a GEI by a generic warping field across various speeds does not work well, because the GEI itself is highly affected by the speed changes, which badly affects the generation of the generic warping field. A morphed key-frame better reduces the residuals than a morphed GEI in this visualization example, the accuracy is, however, still the worst of the three morphed features because of its low stability. As a result, only the morphed SSGEI achieves the true match here because of its good trade-off between speed invariance and stability as well as the generic warping field.

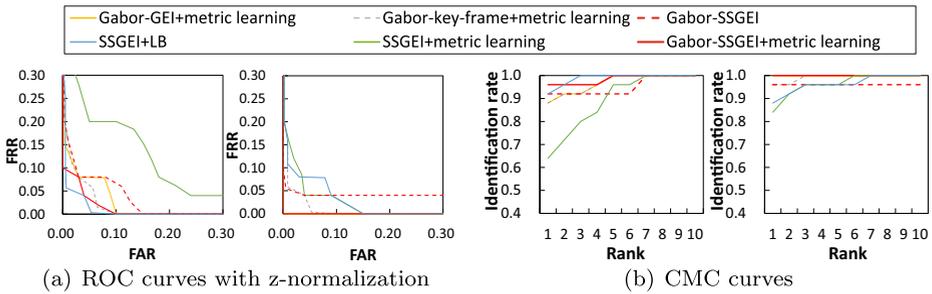


Fig. 8 ROC and CMC curves in OUTD-A for within-walking and within-running cases (left: 7 km/h gallery versus 2 km/h probe, right: 8 km/h gallery versus 10 km/h probe) to analyze individual component contributions

5.6 Contributions of individual components

To confirm the contributions of the individual components of the proposed method, we compared the proposed method with methods excluding individual components: SSGEI + metric learning (excluding Gabor filtering), Gabor-GEI + metric learning and Gabor-key-frame + metric learning (both excluding SSGEI), and Gabor-SSGEI (excluding metric learning). We further combined the proposed SSGEI with the state-of-the-art deep learning-based method, i.e., Local @ Bottom (LB) [45], to compare its contribution with that of traditional metric learning employed in the proposed method. Considering the limited training samples in OUTD-A, we fine-tuned a pre-trained model of LB on the OU-ISIR Large Population Dataset (OULP) [10], which is one of the existing largest gait datasets containing over 4,000 subjects with view variations. Data augmentation for training samples was not applied, because the performance improvement is not obvious compared with the network trained without data augmentation, as reported in [45]. To make a fair comparison, we fine-tuned different network for within-walking, within-running and cross-mode, respectively. In the testing stage, gait mode classification was first applied to the pair of probe and gallery SSGEI, which were then fed into the corresponding network according to the classification results.

We report the accuracies of the above methods for OUTD-A with receiver operating characteristic (ROC) curves with z-normalization and cumulative matching characteristics (CMC) curves. While the ROC curve shows a trade-off between FAR and FRR when an acceptance threshold changes, the CMC indicates rates at which the genuine subjects are included within each rank. We also report the rank-1 identification rate for the CMC curve, and EER with and without z-normalization from the ROC curve.

First, we show the accuracies in the within-walking and within-running cases. The ROC curves with z-normalization and the CMC curves for two pairs of speeds, i.e., 7 km/h gallery versus 2 km/h probe for the within-walking case and 8 km/h gallery versus 10 km/h probe for the within-running case, are shown in Fig. 8. The overall accuracies for all speed combinations of the within-walking and within-running cases are also provided in Table 3(a). In the within-walking case, the proposed method yields the best performance in both identification and verification scenarios, which indicates that the individual components substantially contribute to the proposed method. Under the within-running case, the proposed method achieves the best accuracy as a whole and yields the second-best for EER with z-normalization, which is still a sufficiently low error (0.4%). The deep learning-based

Table 3 Overall rank-1, z-EER, and EER[%] for all speed combinations for the three cases in OUTD-A to analyze the individual component contributions. Metric learning are denoted as ML

(a) Within-walking and within-running cases

	Within-walking			Within-running		
	Rank-1	z-EER	EER	Rank-1	z-EER	EER
SSGEI + ML	87.7	5.2	8.9	95.1	2.7	5.2
Gabor-GEI + ML	96.8	2.4	4.4	100	0.07	4.0
Gabor-key-frame + ML	97.1	2.6	5.7	97.8	1.8	6.7
Gabor-SSGEI	95.1	2.7	6.8	96.9	3.1	5.3
SSGEI + LB	93.8	2.4	3.9	97.3	2.2	3.1
Gabor-SSGEI + ML	99.3	1.3	3.1	100	0.4	3.0

(proposed)

(b) Cross-mode case

Algorithm	Rank-1	z-EER	EER
SSGEI + morphing + ML	54.0	15.6	21.7
Gabor-GEI + morphing + ML	71.2	9.6	17.7
Gabor-key-frame + morphing + ML	77.0	9.1	14.9
Gabor-SSGEI + morphing	65.8	11.4	20.9
SSGEI + morphing + LB	71.3	8.6	12.1
Gabor-SSGEI + ML	58.1	15.3	24.8
Gabor-SSGEI + morphing (w/o AF) + ML	77.3	8.0	13.9
Gabor-SSGEI + morphing + ML (proposed)	81.0	6.9	12.9

framework does, however, not improve the performance compared with the traditional yet effective metric learning method. This is understandable because the dataset we used is quite small although it contains the largest speed variations, which easily leads to the overfitting problem for deep learning models. Nonetheless, the LB still achieves competitive results in the verification scenarios.

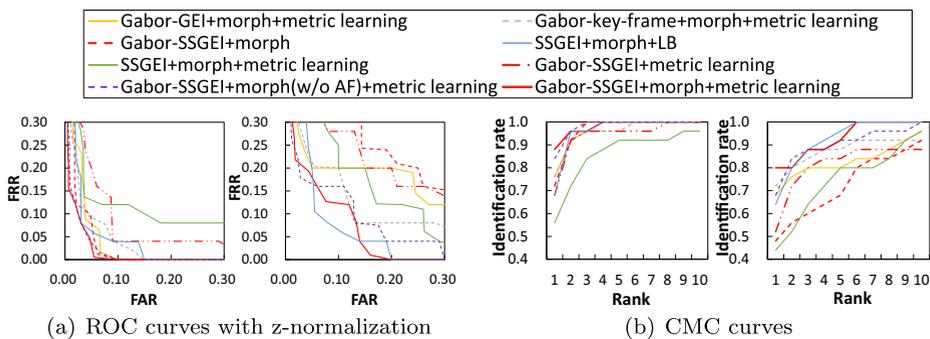


Fig. 9 ROC and CMC curves in OUTD-A for the cross-mode case (left: 8 km/h running gallery versus 7 km/h walking probe, right: 10 km/h running gallery versus 2 km/h walking probe) to analyze individual component contributions

Table 4 Rank-1 identification rates [%] of DCM [17] (before slash) and the proposed method (after slash) for each combination of walking speeds fs , fn , and fq on CASIA-C

P \ G	G		
	fs	fn	fq
fs	96 / 99.8 ± 0.40	92 / 99.5 ± 0.61	88 / 92.8 ± 2.13
fn	88 / 96.5 ± 1.89	97 / 99.8 ± 0.40	89 / 96.0 ± 1.60
fq	89 / 94.4 ± 1.50	93 / 99.5 ± 0.53	95 / 100 ± 0.00

Probe and gallery are denoted as P and G, respectively. For the proposed method, average rank-1 identification rates with standard deviation [%] over 10 random subsets are shown

Next, we evaluate the accuracies of the cross-mode case. Because the morphing process is an additional important component for the cross-mode case, we added the morphing to the above benchmarks and also prepared Gabor-SSGEI + metric learning (excluding morphing) as another benchmark. To demonstrate the effectiveness of the attenuation field, we also report the results of Gabor-SSGEI + morphing (w/o AF) + metric learning (excluding attenuation field from the morphing procedure). The ROC curves with z-normalization and the CMC curves for both pairs of speeds in the cross-mode case, i.e., 8 km/h running gallery versus 7 km/h walking probe and 10 km/h running gallery versus 2 km/h walking probe, are shown in Fig. 9, while the overall accuracies for all speed combinations of the cross-mode case are provided in Table 3(b). As a result, the proposed method still achieves the best overall performance similarly to the within-walking and within-running cases, and Gabor-SSGEI + morphing (w/o AF) + metric learning yields the second-best performance, which shows the mitigation of subject-dependent residuals using the attenuation field is necessary for the proposed generic warping between walking and running modes. In contrast, if we exclude the morphing component, the rank-1 identification rate as well as the EER with and without z-normalization significantly drops below the best ones, 81.0%, 6.9%, and 12.9% for the proposed method, to 58.1%, 15.3%, and 24.8% for Gabor-SSGEI + metric learning (excluding morphing), respectively, which demonstrates that not only the SSGEI, Gabor-filtering, and metric learning components, but also the additional morphing component make considerable contributions to the high accuracy of the proposed method.

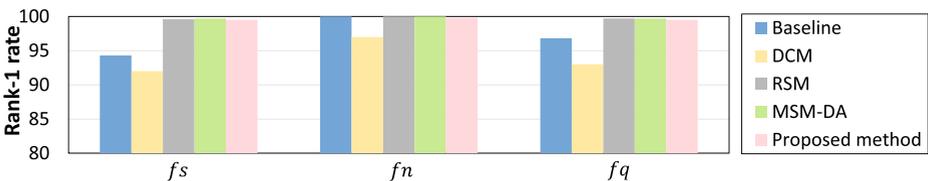
**Fig. 10** Rank-1 identification rates [%] of benchmark algorithms for gallery speed fn on CASIA-C. Here, fs , fn , and fq represent the probe speeds. Note that RSM [5] uses all 153 subjects for the test set because it does not require auxiliary training subjects, while the other four methods use 120 subjects for the testing set

Table 5 Rank-1 identification rates [%] of the benchmark algorithms for small (3 and 4 km/h) and large (2 and 6 km/h) speed changes on OUTD-A

Speed change	HMM [23]	SN [41]	STM [27]	DCM [17]	RSM [5]	MSM [12]	MSM-DA [11]	LB-sep [45]	LB-uni [45]	Ours
Small	84	–	90	98	100	100	100	98	98	100
Large	–	35	58	82	95	84	100	92	88	98

5.7 Comparison with state-of-the-art methods

5.7.1 CASIA-C

In this section, the proposed method is compared with three latest benchmark methods of speed-invariant gait recognition which provided the results on CASIA-C, i.e., DCM [17], RSM [5] and mutual subspace method using divided area (MSM-DA) [11]. We first compare the rank-1 identification rates of each combination of three walking speeds fs , fn , and fq with DCM, which also used 120 subjects for the testing set, in Table 4. The proposed method clearly outperforms DCM for all combinations, particularly for large speed changes (e.g., fq versus fs). Next, following the experimental protocol in [5], we evaluated on pairs of gallery fn versus probe fs , fn , and fq to compare the results with DCM, RSM and MSM-DA as well as the baseline (i.e., simply using GEI), as shown in Fig. 10³. Although it is difficult to make a fair comparison between RSM and the other benchmarks because of the slight difference in gallery size, RSM, MSM-DA and the proposed method achieve almost saturated accuracies (approximately 100% rank-1 identification rate) for all cases.

5.7.2 OUTD-A

In this section, the proposed method is compared with additional state-of-the-art methods of speed-invariant gait recognition, i.e., the hidden Markov model (HMM)-based approach [23], stride normalization (SN) [41], speed transformation model (STM) [27], DCM [17], RSM [5], MSM [12], MSM-DA [11], and the state-of-the-art deep learning-based method, i.e., LB [45] using GEI on OUTD-A. For LB, we applied two strategies, i.e., separately fine-tuned different models for within-walking, within-running, and cross-mode cases (LB-sep), and fine-tuned a unified model using all the GEIs regardless of the gait mode variations (LB-uni). Although some of the benchmarks employed different data sets, the number of subjects and speed difference are almost consistent with those in OUTD-A and hence we also follow the same setting as suggested in [5, 27] to make a comparison that is as fair as possible.

More specifically, HMM was evaluated with a different gait data set whose walking speed pair are 3.3 km/h and 4.5 km/h, and hence the other methods were compared using the matching results between 3 km/h and 4 km/h. To compare with SN, which also employed

³Settings of the number of sequences per subject in the gallery and probe are not clarified in the literature of DCM, while the baseline, RSM, MSM-DA and proposed method use the same settings, i.e., the gallery contains three fn sequences and probes contain the remaining one fn , two fs , and two fq sequences per subject.

Table 6 Rank-1 identification rates [%] of the proposed method for all 81 combinations of speeds on OUTD-A

P \ G	G								
	2km/h	3km/h	4km/h	5km/h	6km/h	7km/h	8km/h	9km/h	10km/h
2km/h	100	100	100	100	96	96	80	64	80
3km/h	100	100	100	100	100	92	84	68	76
4km/h	100	100	100	100	100	92	92	72	80
5km/h	100	100	100	100	100	100	88	80	84
6km/h	100	100	100	100	100	100	92	88	84
7km/h	100	100	100	100	100	100	88	84	84
8km/h	76	80	76	96	84	80	100	100	100
9km/h	76	80	80	84	84	80	100	100	100
10km/h	68	80	80	84	80	80	100	100	100

Probe and gallery are denoted as P and G, respectively

a different gait data set whose walking speed pair are 2.5 km/h and 5.8 km/h, we chose the matching results between 2 km/h and 6 km/h for the other methods.

Results are shown in Table 5. In addition, Table 7⁴ listed the rank-1 identification rates averaged over all combinations of speeds in the within-walking, within-running, and cross-mode cases for last seven methods in Table 5. Because RSM only provided results for the cross-mode of gallery speeds from 2 km/h to 7 km/h while the probe speeds are from 8 km/h to 10 km/h, the averaged rank-1 identification rate of cross-mode are computed over these 36 combinations. Moreover, the rank-1 identification rates of 81 combinations of all walking and running speeds for the proposed method are reported in Table 6.

In Tables 5–7, the proposed method achieves the second-best performance in the within-walking case, which is competitive with the best one (i.e., MSM-DA) considering the small number of test subjects in this dataset (i.e., 25 subjects). Although MSM-DA obtains the highest accuracies in the walking case via focusing on static parts that are less affected by walking speed variations, we point out that this method is unsuitable to be extended to the cross-mode case, where both static parts and dynamic parts vary between the walking and running modes (see Fig. 2). In within-running and cross-mode cases, the proposed method clearly outperforms the other algorithms, which even yields better results than the state-of-the-art deep learning-based method (i.e., LB) by approximately 10% with respect to the averaged rank-1 identification rate for the cross-mode case.

5.8 Evaluation of computational time

To evaluate the computational cost, MATLAB code of the proposed method was run on a PC with an Intel Core i7 4.00 GHz processor and 32 GB RAM. The training time of the generic warping field, the optimization time for the duration parameter p and metric learning, as well as the query time of each sequence are listed in Table 8. Although training the generic warping field takes a relatively long time, this process can be done offline beforehand. We further compare the computation time with RSM [5] in Table 9 for the within-walking

⁴Results of RSM are read from the figures in the original paper.

Table 7 Overall rank-1 identification rates [%] of the proposed method and other benchmarks for all three modes on OUTD-A

Algorithms	Within-walking	Within-running	Cross-mode
DCM [17]	92.4	–	–
RSM [5]	98.1	99.2	53.4
MSM [12]	96.8	–	–
MSM-DA [11]	99.8	–	–
LB-sep [45]	94.3	95.6	58.0
LB-uni [45]	93.3	98.2	70.2
Proposed method	99.3	100	80.4

Table 8 Computation time of the proposed method. Metric learning is denoted as ML

Process	Cross-mode	Within-walking	Within-running
Training of warping field [h]	3.133	–	–
Training of parameter p [s]	0.015	0.009	0.003
Training of ML [s]	0.170	0.115	0.059
Query time per sequence [s]	0.010	0.012	0.033

Table 9 Computation time [s] of the proposed method and RSM [5] in within-walking case

Running stage \ Method	RSM [5]	Proposed method	Proposed method (estimated)
Machine specification	Intel Core i5 3.10 GHz, 16 GB RAM	Intel Core i7 4.00 GHz, 32 GB RAM	75% computational power
Training time	320.09	0.124	0.374
Query time per sequence (#Galleries)	0.60 (122)	0.012 (25)	0.078 (122)

Table 10 Correct classification rates [%] for each subset by classifying the gait mode into two classes and three classes

	Training	Gallery in test	Probe in test
2 classes	100	100	100
3 classes	100	94.71	94.32

Table 11 Overall rank-1, z-EER, and EER [%] of using two gait mode classes and three gait mode classes for within-walking, within-running, and walking vs. running cases

	2 classes			3 classes		
	Rank-1	z-EER	EER	Rank-1	z-EER	EER
Within-walking	99.3	1.3	3.1	98.0	1.3	3.7
Within-running	100	0.4	3.0	100	0.2	2.8
Walking vs. running	81.0	6.9	12.9	81.9	7.9	13.0

case. Because of the different numbers of gallery sequences⁵ and machine specifications, we estimate the proposed method under a comparable setting. The result illustrates that the computational cost of the proposed method is much lower than that of RSM and hence more suitable for real applications.

5.9 Effect of number of gait mode classes

To evaluate the effect of the number of gait mode classes, we tested the performance of the proposed method by classifying the gait mode into three classes, i.e., slow-walking (from 2 km/h to 4 km/h), fast-walking (from 5 km/h to 7 km/h), and running (from 8 km/h to 10 km/h). Considering the effectiveness of the classifier using SSGEI and the gait period T [frames] reported in Section 5.3, we first adopted SSGEI concatenated with the gait period T [frames] to classify the walking and running mode, and then used GEI and the gait period T [frames] for the classification of slow-walking and fast-walking, which appears obvious changes in the dynamic parts affected by the speed variation. The results in Table 10 show that the classification accuracy degrades by using three classes, which is understandable as the difficulty of classification raised with the increase of the number of classes.

The performance in both verification and identification scenarios of using three gait modes and two gait modes are compared in Table 11. For a fair comparison, the results of three gait modes in the within-walking case are computed as an overall performance of within slow-walking, within fast-walking, and slow-walking versus fast-walking, while the results in the case of walking versus running are computed for both slow-walking versus running and fast-walking versus running, respectively. As shown in Table 11, the performance of three gait modes in the within-walking case are worse than two modes, which is mainly caused by the misclassification of gait modes. On the other hand, using three gait modes yields higher identification accuracy in the case of walking versus running, because the finer warping fields of three modes generated better transformation results than a general warping field between walking and running of two gait modes. Therefore, it is a trade-off between fine warping fields and difficulties in the gait mode classification when choosing the appropriate number of gait mode classes.

6 Conclusion

This paper presented a framework for speed-invariant gait recognition using a speed invariant and stable gait representation called SSGEI. To realize a good trade-off between the

⁵The computational cost of RSM was evaluated on USF dataset [34], which contains 122 subjects in the gallery set.

speed invariance and stability, SSGEI is computed by aggregating multiple frames over the optimal duration around single-support phases, which are chosen by maximizing the Fisher ratio using a training set. For the challenging cross-mode case, SSGEI is further morphed into intermediate poses between walking and running using an FFD-based generic warping field across the population as well as an attenuation field based on the trash bin concept to suppress subject-dependent residuals. For better performance, Gabor filters and metric learning are combined with SSGEI as postprocessing steps. Comprehensive experiments using two publicly available gait data sets, CASIA-C and OUTD-A, demonstrated the effectiveness and efficiency of the proposed method.

In this work, we applied the proposed SSGEI only to speed-invariant gait recognition. Because the static part enhancement of the SSGEI may be also effective for other covariates in gait recognition (e.g., the forward-backward arm swing observed from a side view may not be observed from a frontal view), a future direction is to evaluate the accuracy of gait recognition under other covariates using the SSGEI. On the other hand, the static parts may be more affected than the dynamic parts for some covariates such as clothing and carrying status, and another future research avenue is therefore to seek a gait representation that highlights the dynamic parts, in contrast to the proposed SSGEI. Additionally, although the performance under the within-walking and within-running cases seems to be saturated, the cross-mode gait recognition still requires more exploration. Rather than generating generic warping field across the population, we plan to extend it to a subject-dependent deep learning-based framework after sufficient data are collected, which helps to improve the accuracy in the cross-mode scenario.

Acknowledgements This work was supported by JSPS Grants-in-Aid for Scientific Research (A) JP18H04115, by Jiangsu Provincial Science and Technology Support Program (No. BE2014714), by the 111 Project (No. B13022), and by the Priority Academic Program Development of Jiangsu Higher Education Institutions.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Bouchrika I, Goffredo M, Carter J, Nixon M (2011) On using gait in forensic biometrics. *J Forensic Sci* 56(4):882–889
2. Bouchrika I, Nixon M (2008) Exploratory factor analysis of gait recognition. In: Proceedings of the 8th IEEE international conference on automatic face and gesture recognition. Amsterdam, The Netherlands, pp 1–6
3. Cheng F, Christmas WJ, Kittler J (2002) Recognising human running behaviour in sports video sequences. In: Object recognition supported by user interaction for service robots, vol 2, pp 1017–1020
4. Fihl P, Moeslund T (2010) Recognizing Human Gait Types, 183–208 INTECH
5. Guan Y, Li CT (2013) A robust speed-invariant gait recognition system for walker and runner identification. In: Proceedings of the 6th IAPR international conference on biometrics, pp 1–8
6. Han J, Bhanu B (2006) Individual recognition using gait energy image. *IEEE Trans Pattern Anal Mach Intell* 28(2):316–322
7. Hossain MA, Makihara Y, Wang J, Yagi Y (2010) Clothing-invariant gait identification using part-based clothing categorization and adaptive weight control. *Pattern Recogn* 43(6):2281–2291
8. Iosifidis A, Tefas A, Pitas I (2012) Activity-based person identification using fuzzy representation and discriminant learning. *IEEE Trans Inform Forensics Secur* 7(2):530–542. <https://doi.org/10.1109/TIFS.2011.2175921>

9. Iwama H, Muramatsu D, Makihara Y, Yagi Y (2013) Gait verification system for criminal investigation. *IPSIJ Trans Comput Vis Appl* 5:163–175
10. Iwama H, Okumura M, Makihara Y, Yagi Y (2012) The ou-isir gait database comprising the large population dataset and performance evaluation of gait recognition. *IEEE Trans Inform Forensics Secur* 7(5):1511–1521
11. Iwashita Y, Kakeshita M, Sakano H, Kurazume R (2017) Making gait recognition robust to speed changes using mutual subspace method. In: 2017 IEEE International Conference on Robotics and Automation (ICRA), pp 2273–2278
12. Iwashita Y, Sakano H, Kurazume R (2015) Gait recognition robust to speed transition using mutual subspace method. In: *Image analysis and processing — ICIAP 2015: 18th international conference, genoa, italy, september 7-11, 2015, proceedings, Part I*, 141–149. Cham
13. Kim TK, Cipolla R (2009) Canonical correlation analysis of video volume tensors for action categorization and detection. *IEEE Trans Pattern Anal Mach Intell* 31(8):1415–1428. <https://doi.org/10.1109/TPAMI.2008.167>
14. Kusakunniran W, Wu Q, Li H, Zhang J (2009) Automatic gait recognition using weighted binary pattern on video. In: *AVSS '09. Sixth IEEE international conference on Advanced video and signal based surveillance*, pp 49–54
15. Kusakunniran W, Wu Q, Zhang J, Li H (2010) Support vector regression for multi-view gait recognition based on local motion feature selection. In: *Proceedings of IEEE computer society conference on computer vision and pattern recognition 2010*, 1–8, San Francisco CA, USA
16. Kusakunniran W, Wu Q, Zhang J, Li H (2011) Speed-invariant gait recognition based on procrustes shape analysis using higher-order shape configuration. In: *The 18th IEEE International Conference Image Processing*, pp 545–548
17. Kusakunniran W, Wu Q, Zhang J, Li H (2012) Gait recognition across various walking speeds using higher order shape configuration based on a differential composition model. *IEEE Trans Syst Man and Cybern Part B: Cybern* 42(6):1654–1668. <https://doi.org/10.1109/TSMCB.2012.2197823>
18. Lee TS (1996) Image representation using 2d gabor wavelets. *IEEE Trans Pattern Anal Mach Intell* 18(10):959–971. <https://doi.org/10.1109/34.541406>
19. Leow A, Huang SC, Geng A, Becker J, Davis S, Toga A, Thompson P (2005) Inverse consistent mapping in 3d deformable image registration: Its construction and statistical properties. In: *Proceedings of the 19th International Conference on Information Processing in Medical Imaging, IPMI'05*, pp 493–503
20. Li M, Yuan B (2005) 2d-lda: a statistical linear discriminant analysis for image matrix. *Pattern Recogn Lett* 26(5):527–532. <https://doi.org/10.1016/j.patrec.2004.09.007>
21. Li Z, Fu Y, Huang T, Yan S (2008) Real-time human action recognition by luminance field trajectory analysis. In: *Proceedings of the 16th ACM International Conference on Multimedia, MM '08*, pp 671–676
22. Lin G, Milan A, Shen C, Reid I (2017) RefineNet: multi-path refinement networks for high-resolution semantic segmentation. In: *CVPR*
23. Liu Z, Sarkar S (2006) Improved gait recognition by gait dynamics normalization. *IEEE Trans Pattern Anal Mach Intell* 28(6):863–876. <https://doi.org/10.1109/TPAMI.2006.122>
24. Lynnerup N, Larsen P (2014) Gait as evidence. *IET Biometrics* 3(2):47–54. <https://doi.org/10.1049/iet-bmt.2013.0090>
25. Makihara Y, Mannami H, Tsuji A, Hossain M, Sugiura K, Mori A, Yagi Y (2012) The ou-isir gait database comprising the treadmill dataset. *IPSIJ Trans Comput Vis Appl* 4:53–62
26. Makihara Y, Sagawa R, Mukaigawa Y, Echigo T, Yagi Y (2006) Gait recognition using a view transformation model in the frequency domain. In: *Proceedings of the 9th european conference on computer vision. Graz, Austria*, pp 151–163
27. Makihara Y, Tsuji A, Yagi Y (2010) Silhouette transformation based on walking speed for gait identification. In: *Proceedings of the 23rd IEEE conference on computer vision and pattern recognition. San francisco, CA, USA*
28. Makihara Y, Yagi Y (2008) Silhouette extraction based on iterative spatio-temporal local color transformation and graph-cut segmentation. In: *Proceedings of the 19th international conference on pattern recognition. Tampa, Florida USA*
29. Makihara Y, Yagi Y (2010) Earth mover's morphing: Topology-free shape morphing using cluster-based emd flows. In: *Proceedings of the 10th asian conf. on computer vision. Queenstown, New Zealand*, pp 2302–2315
30. Muramatsu D, Shiraishi A, Makihara Y, Uddin MZ, Yagi Y (2015) Gait-based person recognition using arbitrary view transformation model. *IEEE Trans Image Process* 24(1):140–154

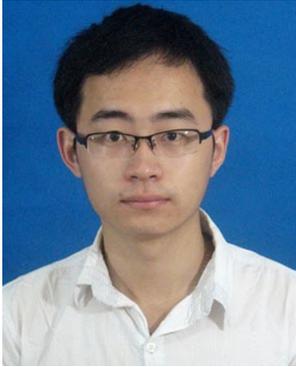
31. Nakajima H, Makihara Y, Hsu H, Mitsugami I, Nakazawa M, Yamazoe H, Habe H, Yagi Y (2012) Point cloud transport. In: Proceedings of the 21st international conference on pattern recognition. Japan, Tsukuba, pp 3803–3806
32. Nixon MS, Tan TN, Chellappa R (2005) Human identification based on gait. Int. Series on biometrics. Springer, Berlin
33. Phillips P, Blackburn D, Bone M, Grother P, Micheals R, Tabassi E (2002) Face recognition vendor test. <http://www.frvt.org>
34. Sarkar S, Phillips J, Liu Z, Vega I, ther PG, Bowyer K (2005) The humanid gait challenge problem: data sets, performance, and analysis. *IEEE Trans Pattern Anal Mach Intell* 27(2):162–177
35. Sederberg TW, Parry SR (1986) Free-form deformation of solid geometric models. *SIGGRAPH Comput Graph* 20(4):151–160. <https://doi.org/10.1145/15886.15903>
36. Shiraga K, Makihara Y, Muramatsu D, Echigo T, Yagi Y (2016) Geinet: view-invariant gait recognition using a convolutional neural network. In: 2016 International Conference on Biometrics (ICB), pp 1–8
37. Takemura N, Makihara Y, Muramatsu D, Echigo T, Yagi Y (2017) On input/output architectures for convolutional neural network-based cross-view gait recognition. *IEEE Trans Circ Syst Video Technol PP(99)*:1–1
38. Tan D, Huang K, Yu S, Tan T (2006) Efficient night gait recognition based on template matching. In: Proceedings of the 18th international conference on pattern recognition, vol 3. Hong Kong, China, pp 1000–1003
39. Tan D, Huang K, Yu S, Tan T (2007) Orthogonal diagonal projections for gait recognition. In: 2007 IEEE International conference on image processing, vol 1, pp i – 337–i – 340
40. Tan D, Huang K, Yu S, Tan T (2007) Uniprojective features for gait recognition. In: Proceedings of the 2007 International Conference on Advances in Biometrics, ICB'07, pp 673–682
41. Tanawongsuwan R, Bobick A (2004) Modelling the effects of walking speed on appearance-based gait recognition. In: Proceedings of the 17th IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol 2, pp 783–790. <https://doi.org/10.1109/CVPR.2004.158>
42. Tao D, Li X, Wu X, Maybank SJ (2007) General tensor discriminant analysis and gabor features for gait recognition. *IEEE Trans Pattern Anal Mach Intell* 29(10):1700–1715. <https://doi.org/10.1109/TPAMI.2007.1096>
43. Wolf T, Babae M, Rigoll G (2016) Multi-view gait recognition using 3d convolutional neural networks. In: 2016 IEEE International Conference on Image Processing (ICIP), pp 4165–4169
44. Wu Z, Huang Y, Wang L (2015) Learning representative deep features for image set analysis. *IEEE Trans Multimed* 17(11):1960–1968. <https://doi.org/10.1109/TMM.2015.2477681>
45. Wu Z, Huang Y, Wang L, Wang X, Tan T (2017) A comprehensive study on cross-view gait based human identification with deep cnns. *IEEE Trans Pattern Anal Mach Intell* 39(2):209–226
46. Xu C, Makihara Y, Li X, Yagi Y, Lu J (2016) 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 2016). Taipei, Taiwan, pp 52–67
47. Xu D, Huang Y, Zeng Z, Xu X (2012) Human gait recognition using patch distribution feature and locality-constrained group sparse representation. *IEEE Trans Image Process* 21(1):316–326. <https://doi.org/10.1109/TIP.2011.2160956>
48. Yam C, Nixon MS, Carter JN (2002) On the relationship of human walking and running: automatic person identification by gait. In: Object recognition supported by user interaction for service robots, vol 1, pp 287–290
49. Yang J, Zhang D, Frangi AF, Yang JY (2004) Two-dimensional pca: a new approach to appearance-based face representation and recognition. *IEEE Trans Pattern Anal Mach Intell* 26(1):131–137. <https://doi.org/10.1109/TPAMI.2004.1261097>
50. Yu H, min Sun G, xing Song W, Li X (2005) Human motion recognition based on neural network. In: Proceedings of the 2005 international conference on communications, circuits and systems, vol 2, p 982
51. Zhang C, Liu W, Ma H, Fu H (2016) Siamese neural network based gait recognition for human identification. In: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp 2832–2836



Chi Xu received the B.S. degree in computer science and technology from Nanjing University of Science and Technology (NUST), China, in 2012. She is currently a Ph.D. candidate in pattern recognition and intelligent system at NUST. Since Jan. 2016, she has been with the Institute of Scientific and Industrial Research, Osaka University, Japan, as a visiting researcher. Her research interests are gait recognition, machine learning and image processing.



Yasushi Makihara received the B.S., M.S., and Ph.D. degrees in Engineering from Osaka University in 2001, 2002, and 2005, respectively. He is currently an Associate Professor of the Institute of Scientific and Industrial Research, Osaka University. His research interests are gait recognition, morphing, and temporal super resolution. He is a member of IPSJ, RJS, and JSME.



Xiang Li received the B.S. degree in computer science and technology in Nanjing University of Science and Technology (NUST) in China in 2012. He is currently a Ph.D. candidate at NUST. Since Jan. 2016, he has been with the Institute of Scientific and Industrial Research, Osaka University as a visiting researcher. His research interests are gait recognition, image processing and machine learning.



Yasushi Yagi is the Executive Vice President of Osaka university. He received his Ph.D. degrees from Osaka University in 1991. In 1985, he joined the Product Development Laboratory, Mitsubishi Electric Corporation, where he worked on robotics and inspections. He became a Research Associate in 1990, a Lecturer in 1993, an Associate Professor in 1996, and a Professor in 2003 at Osaka University. He was also Director of the Institute of Scientific and Industrial Research, Osaka university from 2012 to 2015. International conferences for which he has served as Chair include: FG1998 (Financial Chair), OMINVIS2003 (Organizing chair), ROBO2006 (Program co-chair), ACCV2007 (Program chair), PSVIT2009 (Financial chair), ICRA2009 (Technical Visit Chair), ACCV2009 (General chair), ACPR2011 (Program co-chair) and ACPR2013 (General chair). He has also served as the Editor of IEEE ICRA Conference Editorial Board (2007–2011). He is the Editorial member of IJCV and the Editor-in-Chief of IPSJ Transactions on Computer Vision and Applications. He was awarded ACM VRST2003 Honorable Mention Award, IEEE ROBO2006 Finalist of T.J. Tan Best Paper in Robotics, IEEE ICRA2008 Finalist for Best Vision Paper, MIRU2008 Nagao Award, and PSIVT2010 Best Paper Award. His research interests are computer vision, medical engineering and robotics. He is a fellow of IPSJ and a member of IEICE, RSJ, and IEEE.



Jianfeng Lu received his B.S. degree in computer software and his M.S. degree and Ph.D. in pattern recognition and intelligent system from Nanjing University of Science and Technology (NUST), China. He is currently a Professor in the same school. His research interests include image processing, pattern recognition, and data mining.