

## Gait Recognition under Speed Transition

Al Mansur, Yasushi Makihara, Rasyid Aqmar and Yasushi Yagi  
 The Institute of Scientific and Industrial Research  
 Osaka University, 8-1 Mihogaoka, Ibaraki, Osaka, 567-0047 Japan  
 {mansur,makihara,aqmar,yagi}@am.sanken.osaka-u.ac.jp

### Abstract

*This paper describes a method of gait recognition from image sequences wherein a subject is accelerating or decelerating. As a speed change occurs due to a change of pitch (the first-order derivative of a phase, namely, a gait stance) and/or stride, we model this speed change using a cylindrical manifold whose azimuth and height corresponds to the phase and the stride, respectively. A radial basis function (RBF) interpolation framework is used to learn subject specific mapping matrices for mapping from manifold to image space. Given an input image sequence of speed transitioned gait of a test subject, we estimate the mapping matrix of the test subject as well as the phase and stride sequence using an energy minimization framework considering the following three points: (1) fitness of the synthesized images to the input image sequence as well as to an eigenspace constructed by exemplars of training subjects; (2) smoothness of the phase and the stride sequence; and (3) pitch and stride fitness to the pitch-stride preference model. Using the estimated mapping matrix, we synthesize a constant-speed gait image sequence, and extract a conventional period-based gait feature from it for matching. We conducted experiments using real speed transitioned gait image sequences with 179 subjects and demonstrated the effectiveness of the proposed method.*

### 1. Introduction

Human gait has been recently recognized as one of several biometric modalities for verifying people's identity. It has several unique advantages over other biometric modalities (e.g., DNA, fingerprint, finger and palm veins, iris, ear, and face) because it can be recognized even if a subject is located far from the camera (or the image resolution is low) and is also uncooperative (e.g., a perpetrator captured by a CCTV camera). Gait recognition was used to provide evidence against a burglar in the United Kingdom court [5], and the development of a gait-based identity verification system for criminal investigation was reported in [13].

In spite of considerable progress in gait recognition research, invariance to large intra-class variations of gait features is still a challenging problem (e.g., walking speed [33, 35, 2], viewpoint [23, 15, 27], clothing [12], and elapsed time [31, 29]). Among these features, speed invariance is one of the most important issues because walking speed often varies depending on situation.

Gait recognition from a speed-transitioned gait image sequence is an important problem because a speed-transitioned gait image sequence is commonly observed in many situations where people occasionally decelerate and accelerate. For example, people decelerate in front of a red pedestrian crossing signal and accelerate after the signal turns green. In addition, people decelerate when approaching the door of a room, shop, or building, and accelerate after passing through the doorway. Walking speed change is considered from two aspects: (1) walking speed change between a matching pair of gait image sequences and (2) walking speed change within a gait image sequence. While most studies of speed-invariant gait recognition [33, 35, 2] consider the first aspect as cross-speed gait recognition, so far very limited attention has been paid to the second aspect.

Walking speed change is essentially derived from two factors: (1) the change in pitch (the first-order derivative of phase, namely, gait stance) and (2) the change in stride. If the speed transition is only derived from the first factor, the problem can be solved by non-linear time warping techniques [6, 37, 2]. However, both factors simultaneously change under a real speed transition, and hence rate adjustment alone does not work for the speed-transitioned gait image sequence.

Therefore, in this paper, we tackle this challenging problem: gait recognition from a speed-transitioned gait image sequence. The contributions of this paper are summarized as follows.

**Cylindrical manifold to simultaneously model phase and stride:** We employ a cylindrical manifold whose azimuth and height corresponds to the phase and the stride, respectively, and define a subject-dependent mapping function to map an arbitrary point on the manifold to image space.

**Unified framework of reconstruction-based and exemplar-based methods of estimating subject-dependent mapping function:** Because it is difficult to estimate the subject-dependent mapping function from only a limited observation of a speed-transited gait image sequence, we construct an eigenspace of mapping functions comprising an auxiliary set of the phase-synchronized multi-stride, multi-subject exemplar gait image sequences. Using an energy minimization framework, we then estimate the subject-dependent mapping function based on the fitness to both the input speed-transited gait image sequence and the eigenspace.

**Pitch-stride preference model to robustly estimate the phase and the stride:** Robust estimation of phase and stride is essential for accurate estimation of the subject-dependent mapping function, and hence we introduce a pitch-stride preference model that defines a subject-dependent relationship between pitch and stride under speed variation. Parameters of the pitch-stride preference models are also estimated in the above energy minimization framework.

## 2. Related work

**Gait features:** Gait features used for recognition can be mainly divided into two types: period-based features and frame-based features.

Period-based features are typically computed from one gait period consisting of several frames. Among these period-based features, averaged silhouette [20] (also known as gait energy image (GEI) [11]) is one of the most frequently used features for gait recognition. Other types of period-based gait features include masked GEI [4], frequency-domain features [23], gait entropy images [3], gait flow images [16], chrono-gait images [38], Fourier descriptors [39], and Gabor features [34]. Although these gait features are basically normalized by the gait period, i.e. they are invariant to the change of the gait period, they are not invariant to the pitch change within the gait period.

Frame-based features are computed from individual frames or few consecutive frames. The simplest one is the raw silhouette itself [31, 21, 6] or its projection into principal component (PCA) space [30]. Other types include, but are not limited to, width vector [8] and frieze pattern [19]. Although the frame-based features can handle the pitch change within the gait period in conjunction with non-linear time warping, they are sensitive to silhouette segmentation errors.

In addition, both period-based and frame-based features are significantly affected by change in stride, and hence they require another speed-invariant framework.

**Non-linear time warping for gait recognition:** To cope with the change of pitch within the gait period, many researchers employ dynamic time warping for frame-based raw silhouette sequence [6, 37] or width vector [8].

As a probabilistic representation of non-linear time warping, a hidden Markov model (HMM) is also frequently employed [32, 7]. For example, Aqmar et al. [2] proposed a gait recognition method using a combination of cubic higher-order local auto-correlation features and the HMM. Liu et al. [21] presented a dynamics-normalized gait recognition algorithm using a population HMM (pHMM). Using the Viterbi decoding, a given sequence is mapped onto the pHMM states and therefore normalized. The average of the stances of each state provides a dynamics normalized gait period of fixed length. Although these approaches can cope with the change of pitch within the gait period, they cannot handle the change in stride.

**Adaptation to stride change for gait recognition:** Compared with adaptation to changes in pitch, there are few studies on adaptation to changes in stride for gait recognition. Tanawongsuwan et al. [33] proposed a method of stride normalization of double-support gait silhouettes based on statistical relationship between walking speed and stride. They used only five silhouettes (two single-support images and three double-support images) for recognition and discarded the other informative images. Tsuji et al. [35] proposed a method of gait silhouette transformation from one speed to another to deal with walking speed variation in gait recognition. Based on the assumption that dynamic features are dependent on speed changes while static features are independent of speed changes, the dynamic and static features are extracted separately. Therefore, speed transformation is applied to the dynamic features only. Kusakunniran et al. [14] deployed Procrustes shape analysis to adapt to speed changes, and exploited the higher-order shape configuration as a gait feature. These two methods assume that walking speed is constant within the gait period, and hence may fail to deal with the speed-transited gait image sequence.

**Use of auxiliary set for robust gait recognition:** Another line of research related to our work is gait recognition using an auxiliary set to improve the robustness with respect to a variety of covariates (e.g., viewpoints). Makihara et al. [23] proposed a view transformation model (VTM) via factorization of a matrix composed of an auxiliary set of non-recognition targets. Kusakunniran et al. [15] re-formulated the cross-view gait recognition as a regression problem, and proposed a VTM using support vector regression. Martin-Felez and Xiang [28] cast the gait recognition problem as a bipartite ranking problem using an auxiliary set and demonstrated the effectiveness for any covariate types. All of the above methods adopt period-based gait features, and hence they cannot cope with a speed-transition within a gait period.

**Manifold learning:** For gait and human motion analysis, low-dimensional manifold representation has been used by several researchers. For example, for 3D person tracking,

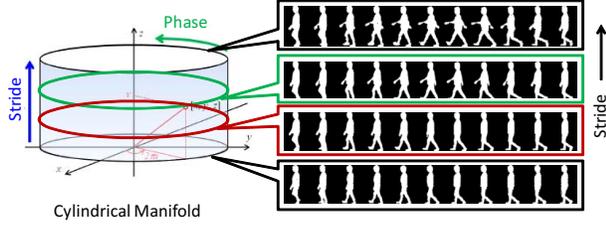


Figure 1. Phase-synchronized multi-stride exemplars and their location on the cylindrical manifold.

Lee et al. [17] combined kinematics and visual manifolds and Urtasun et al. [36] used a scaled Gaussian process latent variable model to learn prior models. In [9], Elgammal et al. used a nonlinear manifold to separate the style and content of human gait; Lee et al. [18] used cylindrical manifold embedding for hand rotation and grasp tracking from a single IR camera; and in a person identification task [7], a Gaussian process latent variable model was used for low-dimensional embedding and an HMM was used for modeling the temporal dynamics. Most of these works mainly deal with the tracking problem, where subject individuality is ignored. However, in our work, instead of tracking, we synthesize the human silhouette sequence at different speeds. This problem formulation requires that the synthesized silhouettes are very close to the observation. In a recent work on view invariant gait recognition [27], a torus-based manifold was used to reconstruct a gait sequence at any desired view. However, in this work, reconstruction was based on the exemplars only, and therefore required a large amount of training data with rich variation of gait styles and shapes. In another work [1], the problem of gait recognition from low frame-rate video is handled. Although both exemplar-based and reconstruction-based cues were used to synthesize a high frame-rate gait sequence, it was assumed that the person was walking at a constant speed.

### 3. Modeling speed transition

#### 3.1. Cylindrical manifold representation

To represent gait with speed variation, a cylindrical manifold (see Fig. 1) is introduced to associate an image with an arbitrary pair of phases and strides (representing speed) where the azimuth angle indicates the phase and the vertical position indicates the stride. A point  $\mathbf{x} \in \mathbb{R}^3$  on the manifold whose phase and stride is  $s$  and  $v$ <sup>1</sup> is then represented by the following mapping function.

$$\mathbf{x} = f(s, v) = [\cos(2\pi s), \sin(2\pi s), v]^T \quad (1)$$

Next, RBF centers on the cylindrical manifold are introduced as  $\{\mathbf{z}_j\} (j = 0, \dots, N^C - 1)$ , which are uniformly

<sup>1</sup> Strictly speaking, as we use height-normalized gait features for recognition in the same way as the previous work, we also use height-normalized stride for this model. For simplicity, we mean height-normalized stride by just stride later.

distributed along azimuth and vertical directions. An arbitrary point  $\mathbf{x}$  on the cylindrical manifold is mapped into the image space as the vector  $\mathbf{y} \in \mathbb{R}^{N^I \times 2}$  by using a radial basis function  $k(\cdot, \cdot)$  between the point and all the RBF centers as

$$\mathbf{y} = D\psi(\mathbf{x}) \quad (2)$$

$$\psi(\mathbf{x}) = [k(\mathbf{x}, \mathbf{z}_0), \dots, k(\mathbf{x}, \mathbf{z}_{N^C-1})]^T \quad (3)$$

$$k(\mathbf{x}, \mathbf{z}_j) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{z}_j\|^2}{2\sigma^2}\right) \quad \forall j, \quad (4)$$

where  $D \in \mathbb{R}^{N^I \times N^C}$  is a mapping matrix from a RBF coefficient vector  $\psi(\mathbf{x}) \in \mathbb{R}^{N^C}$  to the image vector  $\mathbf{y}$  and  $\sigma$  is the bandwidth of RBF.

#### 3.2. Reconstruction-based mapping matrix estimation

The simplest way to estimate the mapping matrix is the direct solution from the observed data. Given a sequence of  $N$  image vectors, phases, and strides as  $Y = [\mathbf{y}_0, \dots, \mathbf{y}_{N-1}] \in \mathbb{R}^{N^I \times N}$ ,  $\mathbf{s} = [s_0, \dots, s_{N-1}]^T$ , and  $\mathbf{v} = [v_0, \dots, v_{N-1}]^T$  for a specific person, the following equation holds,

$$Y = D\Psi(\mathbf{s}, \mathbf{v}), \quad (5)$$

where  $\Psi(\mathbf{s}, \mathbf{v}) = [\psi(f(s_0, v_0)), \dots, \psi(f(s_{N-1}, v_{N-1}))] \in \mathbb{R}^{N^C \times N}$  is a RBF coefficient matrix. Here, if the rank of the RBF coefficient matrix  $\Psi$  is equal to the number of RBF centers  $N^C$ , and  $\mathbf{s}$  and  $\mathbf{v}$  are known, mapping matrix  $D$  can be solved by the following least square

$$\hat{D} = Y\Psi^+, \quad (6)$$

where  $\Psi^+$  is a pseudo-inverse of  $\Psi$  defined as  $\Psi^+ = \Psi^T(\Psi\Psi^T)^{-1}$ . Once the mapping matrix  $D$  is obtained, an image of the specific person with an arbitrary pair of phases and strides is synthesized by Eqs. (1)(2). Note that sequences  $Y', \mathbf{s}', \mathbf{v}'$  for another person gives another mapping matrix  $D'$ , and therefore the mapping matrix contains the walking style information of each person. A limitation of this approach is that  $D$  cannot be solved accurately if limited phase-stride pairs are observed in the test data as the rank of  $\Psi$  is degenerate.

#### 3.3. Example-based mapping matrix estimation

Another method of mapping matrix estimation is using exemplars (gait sequences with set of strides obtained from non-target training subjects (see Fig. 1)). In this approach, we assume that sufficient pairs of phases and strides are observed for those training subjects, and hence mapping matrices are successfully computed using Eq. (6). Thereafter PCA is applied to reduce the dimension, and the mean matrix  $\bar{D}^E$  and  $M$  eigen matrices  $\{D_m^E\} (m = 0, \dots, M - 1)$

<sup>2</sup> A two-dimensional image with size  $N^I$  is unfolded into a one-dimensional vector  $\mathbf{y}$ .

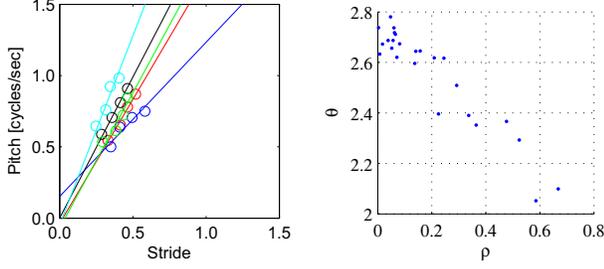


Figure 2. (left) Observed pairs (circles) of strides and pitches under speed variations and pitch-stride preference model (lines) for five training subjects (each color). (right) Distribution of the pitch-stride preference model parameters,  $\rho$  and  $\theta$ .

are obtained as a result. Now, a mapping matrix  $\hat{D}$  for a test subject is approximated by the eigen-space as

$$\hat{D} = \bar{D}^E + \sum_{m=0}^{M-1} a_m D_m^E, \quad (7)$$

where the weight vector  $\mathbf{a} = [a_0, \dots, a_{M-1}]^T$  contains the weights of eigenmatrices  $D_m^E$ . While this exemplar-based method enables us to represent the mapping matrix using exemplars, a major weakness of this approach is that if the training data do not contain sufficient variation of gait styles and shapes, the synthesized images will differ significantly from the observed test data.

### 3.4. Pitch-stride preference model

To estimate a mapping matrix accurately by reconstruction-based and/or exemplar-based approaches, accurate estimation of the phase sequence  $\mathbf{s}$  and the stride sequence  $\mathbf{v}$  is essential. It is, however, difficult to estimate the phase and stride independently frame-by-frame, and it is likely that unstable phase and stride sequence will be estimated.

We notice that each subject has his or her own preferred combination of pitch  $p$  (the first order difference of the phase, expressed as  $p_i = s_i - s_{i-1}$ ) and stride  $v$  under speed variation as shown in Fig. 2(a). In addition, we observe that pairs of the pitches and strides are well represented by a line<sup>3</sup> for each subject. We therefore introduce the pitch-stride preference model as a line in the pitch-stride space to constrain the stride and phase sequences in the estimation step. Specifically, we express the line for the pitch-stride preference model by  $\rho, \theta$  representation as

$$v \cos \theta + p \sin \theta = \rho. \quad (8)$$

Moreover, we introduce prior knowledge of the parameters  $\rho$  and  $\theta$  of the pitch-stride preference model. As shown in Fig. 2(b), the distribution of the pitch-stride preference

<sup>3</sup> Although a polynomial or exponential may fit the data better, due to the limited range of observed pitch and stride, piecewise linearity assumption holds, and straight line representation can be used for simplicity.

model parameters are tightly constrained. We therefore represent this distribution as a single Gaussian distribution  $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ , where  $\boldsymbol{\mu}$  is a two-dimensional mean vector, and  $\Sigma$  is a  $2 \times 2$  covariance matrix. We use this distribution as a prior to constrain the parameters of the pitch-stride preference model of a test subject. Note that the mean vector  $\boldsymbol{\mu}$  and the covariance matrix  $\Sigma$  are computed using the training subjects from the same auxiliary set that was used for learning the eigen space of the mapping matrix.

## 4. Estimation

### 4.1. Energy minimization framework

We formulate the estimation process as an energy minimization framework by considering both reconstruction-based and example-based mapping matrix estimation approaches as well as the pitch-stride preference model. More specifically, an energy function  $S$  is expressed as a sum of three terms: (1) a mapping matrix fitness term  $S_M$ , (2) a pitch and stride smoothness term  $S_S$ , and (3) a pitch-stride preference model term  $S_P$  as

$$S(D, \mathbf{a}, \mathbf{s}, \mathbf{v}, \rho, \theta) = S_M(D, \mathbf{a}, \mathbf{s}, \mathbf{v}) + S_S(\mathbf{s}, \mathbf{v}) + S_P(\rho, \theta, \mathbf{s}, \mathbf{v}) \quad (9)$$

The mapping matrix term  $S_M$  further consists of (1a) a data fitness term: squared errors between the reconstructed images based on the mapping matrix  $D$  and the observed images  $Y$  and (1b) an eigenspace fitness term: squared error between the mapping matrix  $D$  and those represented in the eigenspace as

$$S_M(D, \mathbf{a}, \mathbf{s}, \mathbf{v}) = \lambda_{Md} \|D\Psi(\mathbf{s}, \mathbf{v}) - Y\|_F^2 + \lambda_{Me} \|\bar{D}^E + \sum_{m=0}^{M-1} a_m D_m^E - D\|_F^2, \quad (10)$$

where  $\lambda_{Md}$  and  $\lambda_{Me}$  are the coefficients for individual terms and  $\|\cdot\|_F$  means the Frobenius norm of a matrix.

The pitch and stride smoothness term  $S_S$  consists of the sums of squared second-order derivatives of the pitches and strides as

$$S_S(\mathbf{s}, \mathbf{v}) = \frac{\lambda_{Sp}}{N_{im} - 3} \sum_{i=2}^{N_{im}-2} (p_{i-1} - 2p_i + p_{i+1})^2 + \frac{\lambda_{Sv}}{N_{im} - 2} \sum_{i=1}^{N_{im}-2} (v_{i-1} - 2v_i + v_{i+1})^2, \quad (11)$$

where  $\lambda_{Sp}$  and  $\lambda_{Sv}$  are the coefficients for individual terms, and  $N_{im}$  is the number of input images.

The pitch-stride preference model term  $S_P$  consists of two terms. The first term is the fit of the pitch and stride sequences to the pitch-stride preference model, and the second term is the fit of the pitch-stride preference model pa-

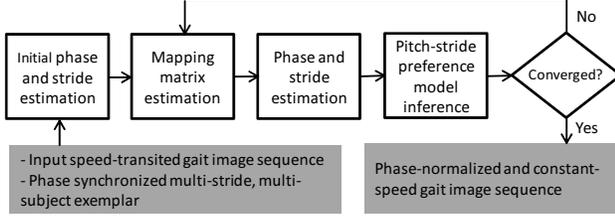


Figure 3. Estimation procedure.

parameter to the prior, expressed as Mahalanobis distance, as

$$S_P(\rho, \theta, \mathbf{s}, \mathbf{v}) = \frac{\lambda_{Pd}}{N_{im}} \sum_{i=0}^{N_{im}-1} (v_i \cos \theta + p_i \sin \theta - \rho)^2 + \lambda_{Pe}([\rho, \theta]^T - \boldsymbol{\mu})^T \Sigma^{-1}([\rho, \theta]^T - \boldsymbol{\mu}) \quad (12)$$

where  $\lambda_{Pd}$  and  $\lambda_{Pe}$  are the coefficients for individual terms.  $v_i$  and  $s_i$  are estimated using dynamic programming (DP) framework (sec. 4.3) and using the definition  $p_i = s_i - s_{i-1}$ . We obtain the optimal mapping matrix  $D^*$ , the weight vector  $\mathbf{a}^*$ , phase sequence  $\mathbf{s}^*$ , stride sequence  $\mathbf{v}^*$ , and pitch-stride preference model parameters  $\rho^*$  and  $\theta^*$  by minimizing the energy function  $S$ .

## 4.2. Iterative solution

The energy function is a quadratic form with regard to the mapping matrix  $D$  and weight vector  $\mathbf{a}$ , and hence can be solved by linear least squares under a fixed phase and stride sequence. On the other hand, the energy function is a nonlinear form with regard to  $\mathbf{s}$ ,  $\mathbf{v}$ ,  $\rho$ , and  $\theta$ , and hence a nonlinear optimizer needs to be introduced to solve it. As the initial solution is crucial for the nonlinear optimizer, DP is used to provide an appropriate initial solution.

For this purpose, we introduce an iterative solution framework shown in Fig. 3. At first, initial phase and stride sequences are estimated by DP as described in the next subsection. Thereafter, the mapping matrix is obtained by linear least squares under the fixed phase and stride sequence. The phase and stride sequence is then solved by a non-linear optimizer, more specifically, the Levenberg-Marquardt method, under the fixed mapping matrix, and subsequently the pitch-stride preference model parameters are solved under the fixed phase and stride sequence. These steps are iterated until the convergence or the maximum number of iterations is reached.

## 4.3. Initial phase and stride estimation

Initial phase and stride sequences are estimated by DP in conjunction with phase-synchronized multi-stride multi-subject exemplars. Let us define a set of quantized phases with equal interval  $\Delta s$  for the DP as  $\{s_k^q | s_k^q = k\Delta s, k = 0, \dots, N^s - 1\}$  and that of quantized strides with equal interval  $\Delta v$  as  $\{v_l^q | v_l^q = v_{min} + l\Delta v, l = 0, \dots, N^v - 1\}$ . In addition, an image vector for the  $n$ -th exemplar subject at

$k$ -th quantized phase and  $l$ -th quantized stride is denoted as  $\mathbf{y}_{k,l}^n$ . Given an input image sequence  $Y = [\mathbf{y}_0, \dots, \mathbf{y}_{N-1}]$ , the optimal correspondence (path) between the input image sequence and each subject of the phase-synchronized multi-stride exemplar image sequence  $\{\mathbf{y}_{k,l}^n\}$  is obtained by the DP, and the optimal subject and its associated path whose cumulative cost is the minimum are adopted.

For this DP scheme, a cumulative cost and the optimal phase and stride transition from the  $(i-1)$ -th frame (previous frame) to the  $i$ -th frame (current frame) at the  $k$ -th quantized phase and  $l$ -th quantized stride in the  $n$ -th exemplar subject, are denoted as  $c^n(i, k, l)$ ,  $p_s^n(i, k, l)$ , and  $p_v^n(i, k, l)$ . Now, the optimal path for the input image sequence is estimated via the following procedure.

1. Initialize cumulative cost  $c^n(0, k, l)$  at the first frame.
$$c^n(0, k, l) = \|\mathbf{y}_{k,l}^n - \mathbf{y}_0\|^2 \quad (13)$$

2. Update transition path  $\{p_s^n(i, k, l), p_v^n(i, k, l)\}$  and cumulative cost  $c^n(i, k, l)$ 

$$\{p_s^n(i, k, l), p_v^n(i, k, l)\} = \arg \min_{t \in R_k^s, u \in R_l^v} c^n(i-1, t, u) \quad (14)$$

$$c^n(i, k, l) = c^n(i-1, p_s^n(i, k, l), p_v^n(i, k, l)) + \|\mathbf{y}_{k,l}^n - \mathbf{y}_i\|^2, \quad (15)$$

where  $R_k^s$  is the set of possible quantized phases in the previous frame, which is transitioned to the  $k$ -th quantized phase in the current frame, and  $R_l^v$  is the set of possible quantized strides in the previous frame, which is transitioned to the  $l$ -th quantized stride in the current frame. They are defined as

$$R_k^s = \{s_t^q | 0 \leq s_k^q - s_t^q \leq s_{tol}\} \quad (16)$$

$$R_l^v = \begin{cases} \{v_u^q | 0 \leq v_l^q - v_u^q \leq v_{tol}\} & \text{(Acceleration)} \\ \{v_u^q | -v_{tol} \leq v_l^q - v_u^q \leq 0\} & \text{(Deceleration)}, \end{cases} \quad (17)$$

where  $s_{tol}$  and  $v_{tol}$  are tolerance parameters to control transition ranges for phase and stride, respectively.

3. Choose the optimal subject  $n^*$  and its associated terminal path  $\{p_s^*(N-1), p_v^*(N-1)\}$ 

$$\{n^*, p_s^*(N-1), p_v^*(N-1)\} = \arg \min_{n,k,l} c^n(N-1, k, l), \quad (18)$$

where  $p_s^*(i)$  and  $p_v^*(i)$  are the optimal quantized phase and stride at the  $i$ -th frame for the optimal subject  $n^*$ .

4. Trace the optimal path backward.
$$p_s^*(i-1) = p_s^{n^*}(i, p_s^*(i), p_v^*(i)) \quad (19)$$

$$p_v^*(i-1) = p_v^{n^*}(i, p_s^*(i), p_v^*(i)) \quad (20)$$

Finally, given the optimal path  $\{p_s^*(i), p_v^*(i)\}$ , the optimal phase and stride sequence is set as  $\{s_i^*, v_i^* | s_i^* = s_{p_s^*(i)}^q, v_i^* = v_{p_v^*(i)}^q\}$ .

## 5. Experiments

### 5.1. Dataset

For the performance evaluation of gait recognition under speed transition, we prepared two datasets: (1) dataset 1 and (2) dataset 2<sup>4</sup>. Although there are several publicly available gait databases with multiple constant speeds (e.g., CMU MoBo Database [10] and the OU-ISIR Gait Database [22]), there are no publicly available gait databases with *speed transition* within an image sequence.

**Dataset 1:** Here the probe set consists of speed transitioned gait sequences recorded from 26 subjects in indoor environment. We placed a poster on the wall and asked each subject to walk toward the poster and gradually decrease walking speed and to finally stop (see the supplementary material). We used the final gait periods of the walking sequences as probe which contain significant change in stride. For gallery set, we collected gait sequences from 179 subjects, which includes the 26 probe subjects, and the subjects walked at constant speed (4 km/h on treadmill) or nearly constant speed (on ground) for few seconds.

**Dataset 2:** For probe set, we collected 25 subjects and each subject walked on the treadmill twice using the automatic speed transition protocol that contained a pair of accelerations from 1 km/h to 5 km/h and decelerations from 5 km/h to 1 km/h. Each acceleration and deceleration was performed within approximately three seconds, and middle subsequences with one second were extracted for acceleration and deceleration. For gallery set, we collected 154 subjects, which includes the 25 probe subjects, and each subject walked at a constant speed (4 km/h) for six seconds.

As an auxiliary training set, we collected 24 separate subjects and each subject walked at 2, 3, 4, and 5 km/h. The phase registration method [24] was applied for each sequence, and the multi-stride phase-normalized image sequences were extracted. A simple frame-shifting scheme was then performed to synchronize the phase among the training subjects and consequently the phase-synchronized multi-speed multi-subject exemplar image sequences were obtained. However, within the same speed sequence, stride differences were observed across the subjects owing to individual pitch-stride preferences. Therefore, we constructed multi-stride multi-subject exemplar image sequences using a shape morphing technique [26]. Both extrapolation and interpolation were used, and the morphing rates were selected so that the strides became well aligned. These multi-stride multi-subject exemplar image sequences were used for learning the eigenspace of the mapping matrix, obtaining a prior for the pitch-stride preference model parameters and the initial phase and stride estimation by DP.

All of the image sequences were captured at 60 fps and binarized by background subtraction-based graph-cut seg-

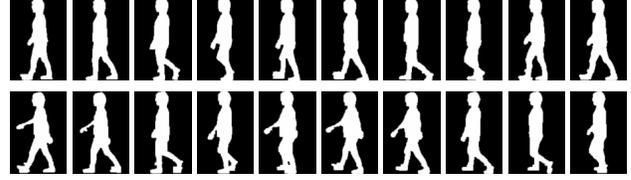


Figure 4. Probe examples of size-normalized image sequences under acceleration (top) and deceleration (bottom) (every 8-th frame from a 60 fps sequence).



Figure 5. Constant-speed reconstruction (bottom) from a decelerated gait image sequence compared with ground truth (top).

mentation [25] and size-normalized into  $22 \times 32$  pixels. Probe examples of size-normalized image sequences under acceleration and deceleration are shown in Fig. 4.

### 5.2. Parameter setting

The number of RBF centers on the cylindrical manifold was set to  $N_s^C = 30$  for azimuth (phase) and  $N_v^C = 4$  for height (stride), which sums up to  $N^C = N_s^C N_v^C = 120$ , and the standard deviation for the RBF kernel was set to  $\sigma = 2\pi/N_s^C$ . To set the coefficients for the energy function, we do a parameter sensitivity analysis (provided in the supplementary material) where we noticed that performance is largely affected by the data term and the exemplar term coefficients, while it is little affected by the other coefficients. The coefficients for the energy function were set as  $\lambda_{Md} = 1$ ,  $\lambda_{Me} = 0.01$ ,  $\lambda_{Sv} = 10^4$ ,  $\lambda_{Sp} = 10$ ,  $\lambda_{Pd} = 10^2$ , and  $\lambda_{Pe} = 0.01$ , respectively. In the DP, the quantization intervals for phase and speed were set as  $\Delta s = 0.01$  and  $\Delta v = 0.01$ , and the tolerance parameters were set as  $s_{tol} = 2.0$  and  $v_{tol} = 1.0$ .

### 5.3. Reconstruction of constant speed gait

In this section, reconstructed constant speed phase-normalized image sequences are compared with original image sequences under the same constant speed for qualitative evaluation as shown in Fig. 5. The results show that the reconstructed phase-normalized image sequence is similar to the constant-speed gallery of the same subject.

### 5.4. Gait recognition

In this section, performance of gait recognition under speed transition is evaluated. The Procrustes shape analysis-based method (denoted as *Procrustes*) [14], HMM-based method (denoted as *HMM*) [2] and Dynamic time warping (DTW) [6] were adopted as benchmarks. The Procrustes shape [14] can handle stride changes to some extent,

<sup>4</sup> <http://www.am.sanken.osaka-u.ac.jp/BiometricDB/GaitST.html>

Performance measure	Acceleration					Deceleration				
	No trans.	Procrustes	HMM	DTW	Proposed	No trans.	Procrustes	HMM	DTW	Proposed
EER (%)	12.0	28.0	9.2	20.0	<b>8.0</b>	12.0	24.0	<b>8.0</b>	19.0	<b>8.0</b>
AUC	0.054	0.217	0.091	0.088	<b>0.014</b>	0.039	0.133	0.067	0.075	<b>0.015</b>
Rank-1 id. rate (%)	48.0	24.0	56.0	40.0	<b>72.0</b>	60.0	36.0	72.0	36.0	<b>84.0</b>
Rank-5 id. rate (%)	76.0	48.0	80.0	52.0	<b>96.0</b>	80.0	60.0	88.0	56.0	<b>92.0</b>

Table 1. EER, AUC, rank-1 and rank-5 identification rates for dataset 2. Bold figures represent the best results.

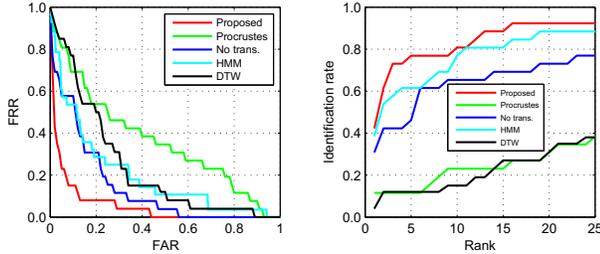
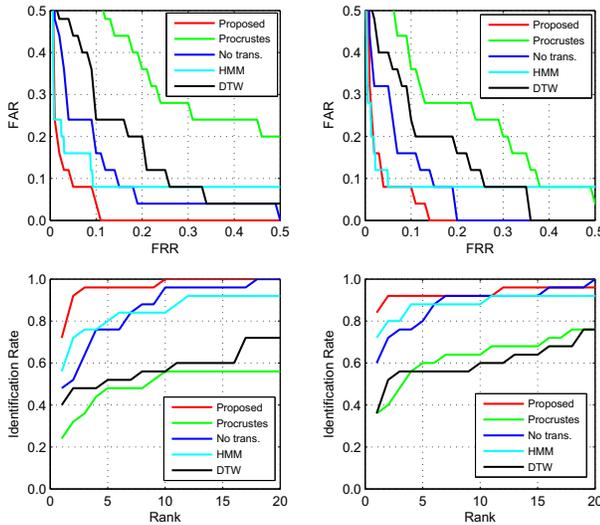


Figure 6. ROC (left) and CMC curves (right) for dataset 1.



(a) Acceleration (b) Deceleration

Figure 7. ROC (top) and CMC curves (bottom) for dataset 2.

while the HMM-based method [2] can handle pitch changes within a gait period. However, DTW considers only one aspect of the speed change: phase evolution speed change, and cannot handle stride change. In addition, direct matching (denoted as *No trans.*) is also considered as a baseline method. GEI [11] was used as gait feature for the proposed method and *No trans.* Two scenarios were considered for the experiments: a verification scenario (one-to-one matching) and an identification scenario (one-to-many matching). In the verification scenario, the receiver operating characteristics (ROC) curve that depicts the trade-off between false acceptance rate (FAR) and false rejection rate (FRR) as well as equal error rate (EER) of the FAR and the FRR, and the area under the ROC curve (AUC) were used as performance measures. As for the identification scenario, the cumulative matching characteristics (CMC) curve itself and also

the rank-1 and rank-5 identification rates picked up from the CMC curve were used as performance measures.

The ROC and CMC curves for dataset 1 and dataset 2 are shown in Fig. 6 and Fig. 7, respectively. Moreover, EER, AUC, rank-1 and rank-5 identification rates dataset 2 are shown in Table 1. The results show that the proposed method outperforms the benchmark methods for both the verification and identification scenarios on both datasets.

In case of dataset 2, it can be noticed that all the methods do better in case of deceleration. The reason is that gait fluctuation (e.g., arm swing, head inclination are different from the gallery) is more evident in case of acceleration. We can also notice that performance of all the methods are worse in dataset 1 compared to those in dataset 2. The reason is that dataset 1 is more challenging as it contains significant gait fluctuation and very short or zero stride lengths (when a person stops). Exemplars used in the proposed method do not contain such variations and therefore we see the performance degradation for the proposed method.

## 6. Conclusion

This paper described a method of gait recognition under speed transition. A cylindrical manifold was introduced to represent an image with an arbitrary phase and stride. While the mapping function from a point on the cylindrical manifold to the image space is estimated by the reconstruction-based and example-based cues, the phase and stride sequences are estimated in conjunction with the pitch-stride preference model. The whole estimation process is formulated in an energy minimization framework with efficient initialization by the DP. The experimental results demonstrated the effectiveness of the proposed method. Future work includes the extension of the proposed method to deal with simultaneous speed and view change, and the formulation of a better pitch-stride model to treat wide range of pitches and strides.

## Acknowledgment

This work was supported by JSPS Grant-in-Aid for Scientific Research (S) 21220003, “R & D Program for Implementation of Anti-Crime and Anti-Terrorism Technologies for a Safe and Secure Society”, Strategic Funds for the Promotion of Science and Technology of the Ministry of Education, Culture, Sports, Science and Technology, the

Japanese Government, and the JST CREST “Behavior Understanding based on Intention-Gait Model” project.

## References

- [1] N. Akae, A. Mansur, Y. Makihara, and Y. Yagi. Video from nearly still: an application to low frame-rate gait recognition. In *Proc. of Computer Vision and Pattern Recognition*, June 2012. [3](#)
- [2] M. R. Aqmar, K. Shinoda, and S. Furui. Robust gait-based person identification against walking speed variations. *IEICE Trans. on Information and Systems*, E95-D:668–676, 2012. [1](#), [2](#), [6](#), [7](#)
- [3] K. Bashir, T. Xiang, and S. Gong. Gait recognition using gait entropy image. In *Proc. of International Conference on Crime Detection and Prevention*, pages 1–6, 2009. [2](#)
- [4] K. Bashir, T. Xiang, and S. Gong. Gait recognition without subject cooperation. *Pattern Recognition Letters*, 31(13):2052–2060, Oct 2010. [2](#)
- [5] I. Bouchrika, M. Goffredo, J. Carter, and M. Nixon. On using gait in forensic biometrics. *Journal of Forensic Sciences*, 56(4):882–889, 2011. [1](#)
- [6] N. V. Boulgouris, K. N. Plataniotis, and D. Hatzinakos. Gait recognition using dynamic time warping. In *Proc. IEEE International Workshop on Multimedia Signal Processing*, pages 263–266, 2004. [1](#), [2](#), [6](#)
- [7] M.-H. Cheng, M.-F. Ho, and C.-L. Huang. Gait analysis for human identification through manifold learning and hmm. *Pattern Recognition*, 41(8):2541–2553, 2008. [2](#), [3](#)
- [8] N. Cuntoor, A. Kale, and R. Chellappa. Combining multiple evidences for gait recognition. In *Proc. of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, volume 3, pages 33–36, 2003. [2](#)
- [9] A. Elgammal and C. S. Lee. Separating style and content on a nonlinear manifold. In *Proc. of Computer Vision and Pattern Recognition*, volume 1, pages 478–485, Washington, DC, USA, Jun. 2004. [3](#)
- [10] R. Gross and J. Shi. The cmu motion of body (mobo) database. Technical report, Robotics Institute, Carnegie Mellon Univ., Jun. 2001. [6](#)
- [11] J. Han and B. Bhanu. Individual recognition using gait energy image. *Trans. on Pattern Analysis and Machine Intelligence*, 28(2):316–322, 2006. [2](#), [7](#)
- [12] M. A. Hossain, Y. Makihara, J. Wang, and Y. Yagi. Clothing-invariant gait identification using part-based clothing categorization and adaptive weight control. *Pattern Recognition*, 43(6):2281–2291, Jun. 2010. [1](#)
- [13] H. Iwama, D. Muramatsu, Y. Makihara, and Y. Yagi. Gait-based person-verification system for forensics. In *Proc. of Int. Conf. on Biometrics: Theory, Applications and Systems*, 2012. [1](#)
- [14] W. Kusakunniran, Q. Wu, J. Zhang, and H. Li. Speed-invariant gait recognition based on procrustes shape analysis using higher order shape configuration. In *Proc. of Int. Conf. on Image Processing*, pages 545–548, 2011. [2](#), [6](#)
- [15] W. Kusakunniran, Q. Wu, J. Zhang, and H. Li. Gait recognition under various viewing angles based on correlated motion regression. *IEEE Trans. Circuits Syst. Video Techn.*, 22(6):966–980, 2012. [1](#), [2](#)
- [16] T. H. W. Lam, K. H. Cheung, and J. N. K. Liu. Gait flow image: A silhouette-based gait representation for human identification. *Pattern Recognition*, 44(4):973–987, 2011. [2](#)
- [17] C. S. Lee and A. Elgammal. Coupled visual and kinematic manifold models for tracking. *International Journal of Computer Vision*, 87(1–2), 2010. [3](#)
- [18] C.-S. Lee and S. W. Park. Tracking hand rotation and grasping from an ir camera using cylindrical manifold embedding. In *Proc. of Int. Conf. on Pattern Recognition*, 2010. [3](#)
- [19] Y. Liu, R. Collins, and Y. Tsin. Gait sequence analysis using frieze patterns. In *Proc. of European Conf. on Computer Vision*, volume 2, pages 657–671, 2002. [2](#)
- [20] Z. Liu and S. Sarkar. Simplest representation yet for gait recognition: Averaged silhouette. In *Proc. of Int. Conf. on Pattern Recognition*, volume 1, pages 211–214, 2004. [2](#)
- [21] Z. Liu and S. Sarkar. Improved gait recognition by gait dynamics normalization. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 28(6):863–876, 2006. [2](#)
- [22] 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 Trans. on Computer Vision and Applications*, 4:53–62, Apr. 2012. [6](#)
- [23] 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 European Conf. on Computer Vision*, pages 151–163, May 2006. [1](#), [2](#)
- [24] Y. Makihara, N. Trung, H. Nagahara, R. Sagawa, Y. Mukaigawa, and Y. Yagi. Phase registration of a single quasi-periodic signal using self dynamic time warping. In *Proc. of Asian Conf. on Computer Vision*, pages 1965–1975, Nov. 2010. [6](#)
- [25] Y. Makihara and Y. Yagi. Silhouette extraction based on iterative spatio-temporal local color transformation and graph-cut segmentation. In *Proc. of Int. Conf. on Pattern Recognition*, Dec. 2008. [6](#)
- [26] Y. Makihara and Y. Yagi. Earth mover’s morphing: Topology-free shape morphing using cluster-based emd flows. In *Proc. of Asian Conf. on Computer Vision*, pages 2302–2315, Queenstown, New Zealand, Nov. 2010. [6](#)
- [27] A. Mansur, Y. Makihara, and Y. Yagi. View-invariant gait recognition from low frame-rate videos. In *Proc. of Int. Conf. on Pattern Recognition*, 2012. [1](#), [3](#)
- [28] R. Martín-Félez and T. Xiang. Gait recognition by ranking. In *Proc. of European Conf. on Computer Vision*, pages 328–341, 2012. [2](#)
- [29] D. Matovski, M. Nixon, S. Mahmoodi, and J. Carter. The effect of time on the performance of gait biometrics. In *Proc. of Int. Conf. on Biometrics: Theory Applications and Systems*, pages 1–6, Sep. 2010. [1](#)
- [30] H. Murase and R. Sakai. Moving object recognition in eigenspace representation: Gait analysis and lip reading. *Pattern Recognition Letters*, 17:155–162, 1996. [2](#)
- [31] S. Sarkar, J. Phillips, Z. Liu, I. Vega, P. Grother, and K. Bowyer. The human id gait challenge problem: Data sets, performance, and analysis. *Trans. on Pattern Analysis and Machine Intelligence*, 27(2):162–177, 2005. [1](#), [2](#)
- [32] A. Sunderesan, A. Chowdhury, and R. Chellappa. A hidden markov model based framework for recognition of humans from gait sequences. In *Proc. IEEE Int’l Conf. on Image Processing 2003*, volume 2, pages 93–96, 2003. [2](#)
- [33] R. Tanawongsuwan and A. Bobick. Modelling the effects of walking speed on appearance-based gait recognition. In *Proc. of Computer vision and pattern recognition*, pages 783–790, 2004. [1](#), [2](#)
- [34] D. Tao, X. Li, X. Wu, and S. Maybank. Human carrying status in visual surveillance. In *Proc. of Computer Vision and Pattern Recognition*, volume 2, pages 1670–1677, Jun. 2006. [2](#)
- [35] A. Tsuji, Y. Makihara, and Y. Yagi. Silhouette transformation based on walking speed for gait identification. In *Proc. of Computer Vision and Pattern Recognition*, Jun 2010. [1](#), [2](#)
- [36] R. Urtaşun, D. J. Fleet, A. Hertzmann, and P. Fua. Priors for people tracking from small training sets. In *Proc. of Int. Conf. on Computer Vision*, pages 403–410, Oct. 2005. [3](#)
- [37] A. Veeraghavan, A. Srivastava, A. K. Roy-Chowdhury, and R. Chellappa. Rate-invariant recognition of humans and their activities. *Trans. Img. Proc.*, 18(6):1326–1339, June 2009. [1](#), [2](#)
- [38] C. Wang, J. Zhang, L. Wang, J. Pu, and X. Yuan. Human identification using temporal information preserving gait template. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2164–2176, 2012. [2](#)
- [39] G. Zhao, R. Chen, G. Chen, and H. Li. Recognition of human periodic movements from unstructured information using a motion-based frequency domain approach. *Image and Vision Computing*, 24:795–809, 2006. [2](#)