Silhouette transformation based on walking speed for gait identification
— **Source link** ↗

Akira Tsuji, Yasushi Makihara, Yasushi Yagi

Institutions: Osaka University

Related papers:

- Individual recognition using gait energy image

- Gait recognition using a view transformation model in the frequency domain

- The humanID gait challenge problem: data sets, performance, and analysis

- Clothing-invariant gait identification using part-based clothing categorization and adaptive weight control

- Improved gait recognition by gait dynamics normalization

Silhouette Transformation based on Walking Speed for Gait Identification

Akira Tsuji, Yasushi Makihara, Yasushi Yagi
Osaka University
8-1 Mihogaoka, Ibaraki, Osaka, 567-0047, JAPAN
{a-tuji, makihara, yagi}@am.sanken.osaka-u.ac.jp

Abstract

We propose a method of gait silhouette transformation from one speed to another to cope with walking speed changes in gait identification. When a person changes his/her walking speed, dynamic features (e.g. stride and joint angle) are changed while static features (e.g. thigh and shin lengths) are unchanged. Based on the fact, firstly, static and dynamic features are separated from gait silhouettes by fitting a human model. Secondly, a factorization-based speed transformation model for the dynamic features is created using a training set for multiple persons on multiple speeds. This model can transform the dynamic features from a reference speed to another arbitrary speed. Finally, silhouettes are restored by combining the unchanged static features and the transformed dynamic features. Evaluation by gait identification using silhouette-based frequency-domain features shows the effectiveness of the proposed method.

1. Introduction

In modern society, there is a growing need to identify individuals in many different situations, including surveillance and access control. For personal identification, many biometric-based authentication methods have been proposed using a wide variety of cues, such as fingerprints, iris, face, and gait. Of these, gait identification has attracted considerable attention because it offers surveillance systems the ability to ascertain identity at a distance.

Various approaches of gait identification have been proposed as model-based approaches[20, 19, 16, 15] and appearance-based approaches[11, 6, 7, 1, 17], and they have shown the effectiveness of gait as a biometrics by using common gait databases including various covariates (e.g. views, shoes, surfaces, clothing, carriages, and walking speed[13][4][5].

Among these covariates, walking speed is one of the most important factor because people often change their walking speed depending on the situation and gait features

such as gait period, arm swing, and stride change largely as a result. While these changes are useful for gait style classification (e.g. walking, running, and jogging)[2], they are just nuisance in the context of gait identification. Hence, to cope with the speed changes, gait feature transformation between different speeds or speed-invariant gait feature extraction is needed.

Liu et al.[8] proposed a HMM-based time-normalized gait feature extraction with standard gait poses and tested it on the slow and fast walking data in the CMU MoBo dataset[4] (3.3 km/h and 4.5 km/h on average respectively). The method does, however, not consider spatial changes (e.g. stride changes) and moreover the range and the number of speed changes used in the experiment are not sufficient. Tanawongsuwan et al. [14] proposed a stride normalization of double-support gait silhouettes based on a statistical relation between the walking speed and the stride. They used only five silhouette (two single-support images and three double-support images) for recognition and discard the other still informative images.

Revisiting the other covariates on gait identification, a factorization-based transformation model on appearance-based gait features was effectively used to adapt to view changes[9]. While view changes are regarded just as appearance changes, speed changes should be treated as changes of dynamic features (e.g. stride) under the condition of keeping static features (e.g. thigh and shin lengths) unchanged.

This paper proposes a method of silhouette transformation between different speeds for gait identification. Unlike the appearance-based factorization, the static and dynamic features are separated by fitting a human model to the silhouette sequence at first, and then factorization is applied only to the dynamic features to transform them from a reference speed to a target speed, while the static features are unchanged. Finally, the unchanged static features and the transformed dynamic features are combined to construct the silhouette sequence in the target speed. Note that the proposed method can deal with not only temporal change (e.g. gait period) but also spatial change (e.g. stride) and that a
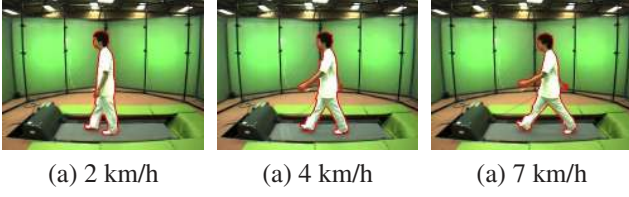
| (a) 2 km/h | (a) 4 km/h | (a) 7 km/h |

Figure 1. Input image (Red countour: segmentation result by [10])

whole sequence of gait silhouette are available for recognition unlike [8][14]. Moreover, at the recognition phase, the proposed silhouette transformation enables us to use not only model-based features but also appearance-based features (e.g. frequency-domain feature[9], frieze pattern[6]), which contain rich information for gait identification.

2. Gait dataset

Because the speed variation of existing gait databases[13][4][14] is insufficient for extensive analysis, we constructed our own speed-variation gait database as *the OU-ISIR Gait Database, Treadmill Dataset A*[18]. Image sequences of walking subjects on a speed-controllable treadmill are captured from side view as shown in Fig. 1. The image size is 640 pixel $\times$ 480 pixel, and the frame rate is 60 fps. The number of subjects is 34 (26 male and 8 female), and speed variation ranges from 2 km/h to 10 km/h at 1 km/h interval. At each speed, a pair of sequences of more than three gait periods were captured. Note that the gait period consists of a pair of left and right steps. In this paper, we focus on speed changes during walking (from 2 km/h to 7 km/h).

As a preprocessing, gait silhouettes are then extracted by background subtraction-based graph-cut segmentation[10], and they are normalized by the height and registered by the region center so as to keep the aspect ratio. We call a series of them Gait Silhouette Volume (GSV).

3. Silhouette transformation

3.1. Overview

Silhouette transformation from a reference speed to a target speed is done by the following three procedures.

1. Static and dynamic features are separately extracted by fitting a human model to the GSV of the reference speed

2. The dynamic features of the reference speed are transformed to those of the target speed by using a pre-trained speed transformation model.

3. The GSV of the target speed is reconstructed by combining the unchanged static features and the transformed dynamic features.

Because speed change makes a large impact on leg motion, in particular, on stride, we focus on transformation of the lower body part in this paper. Details of the procedures are described in the following subsections.

3.2. Separation of static and dynamic features by human model fitting

Human model fitting has been one of central topics in computer vision and a variety of human models (e.g. stick, rectangle, ellipsoid, and generic-cylinder link models) have been proposed[3]. Among these, we adopt a 2-D trapezoidal link model, which can be seen as an approximation of a projected 3-D generic cylindrical link model. Each leg in the model consists of two trapezoids (thigh and shin), a circle (knee joint), and a rectangle (foot). The model is represented by two types of variables: time-dependent variables (dynamic features) and time-independent variables (static feature).

The time-dependent variables of ith leg at tth frame consists of the 2-D waste position $\boldsymbol{W}_i^t$, a joint angle of thigh $\theta_{i,thigh}^t$ and that of shin $\theta_{i,shin}^t$. A joint angle between shin and foot is fixed to be orthogonal in this paper. Thus, the number of variables per leg and frame is 4. Given the number of frames in a gait silhouette sequence as N_{img}, the total nubmer of the time-dependent variables is $8N_{img}$.

On the other hand, the time-independent variables consist of the upper and lower bases, and the height of trapezoids. We assume that the foot size changes in proportion to shin size, in other words, the foot size is dependent variable to the shin size. In addition, note that the lower base of the thigh, the upper base of the shin, and the diameter of knee joint circle are the same, thus the number of the variables per leg is 5 and the total number of them is 10.

Therefore, the total number of the both time-dependent and time-independent variables is $(8N_{img} + 10)$.

Next, the model is fit to the GSV by energy minimization under certain constraints on joint angle range. The energy function is defined by data term related to consistency of the model region and the GSV, and smoothness term to suppress the rapid change of the waste positions and joint angular velocities of the thighs and shins. The energy function is formulated as

$$S = \sum_{t=1}^{N_{img}} (S_{data}(t) + S_{smooth}(t)) \qquad (1)$$

$$S_{data}(t) = w_{out}C_{out}(t) + w_{in}C_{in}(t) \qquad (2)$$

$$S_{smooth}(t) = \sum_{i \in \{left,right\}} \{w_{waist}||\boldsymbol{W}_i^{t+1} - \boldsymbol{W}_i^t||$$
$$+ \sum_{j \in \{thigh,shin\}} w_{angle}(\theta_{i,j}^{t+1} + \theta_{i,j}^{t-1} - 2\theta_{i,j}^t)^2\} \qquad (3)$$
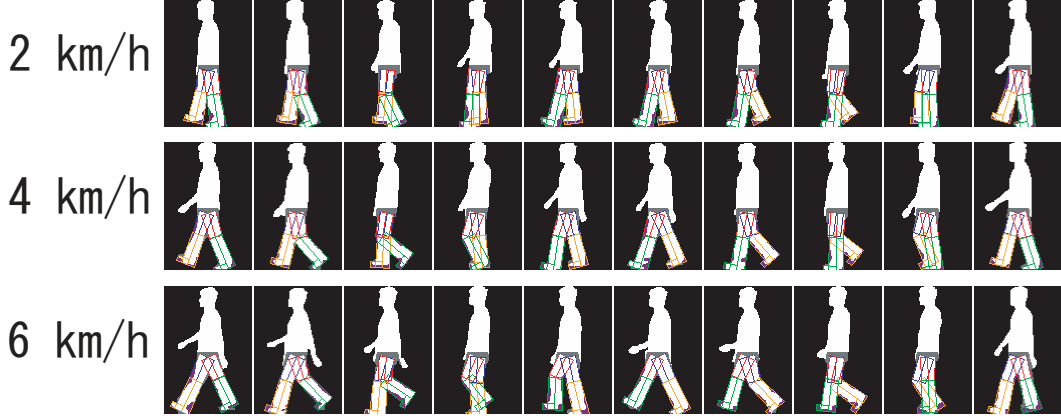
Figure 2. Model fitting results

where $C_{out}(t)$ and $C_{int}(t)$ are areas of silhouette outside the model and non-silhouette inside the model at tth frame respectively, and w_{out} and w_{in} are weights for each term. In addition, w_{waist} and w_{angle} are weights for the temporal changes of the waist positions and those of joint angular velocities. The energy function is minimized by the steepest descent method. The result of model fitting is shown in Fig. 2.

3.3. Dynamic feature transformation

In this section, a factorization-based transformation of the dynamic features is described. Though dynamic features contain the waist positions and joint angles, transformation target is limited to the joint angles for simplicity because significant stride changes are induced by joint angle changes

Because the speed S is generally a function of a gait period T_{period} and a stride L ($S(T_{period}, L) = 2L/T_{period}$ in case of a constant gait), the speed change is divided into the gait period change and the stride change. In addition, if the gait period is regarded as a unit of a feature vector in transformation and matching processes, the feature vector need to be arranged so as that their dimension and gait pose in each dimension can be the same before transformation and matching. Based on the facts, we apply two-step transformation procedures: (1) time normalization and synchronization, and (2) stride transformation.

Time normalization and synchronization First, the gait period N_{gait} is detected as the optimal time shift which maximizes normalized autocorrelation of the GSV for the temporal axis[9]. Then, in order to synchronize gait pose of the feature vector, a double support-phase frame is detected as a key frame where the both legs are the most open, that is, the second moment of the silhouette around the central vertical axis is maximized. Then, the double support frame is set as the first component of the feature vector and then

a pre-defined number N_{nrm} of frames of dynamic features are re-sampled from N_{gait} frames of the original dynamic features. As a result, we obtain the normalized feature vector whose components are synchronous in terms of gait pose and whose dimension is $N_{dim} = 4N_{nrm}$ for the four joint angles.

Formulation of factorization-based stride transformation First, overview of the factorization-based transformation is described as follows. In the training phase, the dynamic features of multiple subjects of multiple strides are collected and a transformation matrix of the dynamic features between different strides is learned. In test phase, given a dynamic feature in a gallery with stride L_{ref} and that in a probe with a stride L, the dynamic feature in the gallery is transformed to that with the stride L. Note that the stride can be estimated from model fitting result for the double support phase.

Next, the factorization-based transformation is formulated as follows. Let the numbers of training subjects and training strides be M and I respectively, and the N_{dim} dimensional feature vector of mth subject of ith stride L_i be $\boldsymbol{a}_{L_i}^m$. Then, we construct a training matrix whose row and column indicates stride and subject respectively and decompose it by singular value decomposition as

$$\begin{bmatrix} \boldsymbol{a}_{L_1}^1 & \cdots & \boldsymbol{a}_{L_1}^M \\ \vdots & \ddots & \vdots \\ \boldsymbol{a}_{L_I}^1 & \cdots & \boldsymbol{a}_{L_I}^M \end{bmatrix} = USV^T = \begin{bmatrix} P_{L_1} \\ \vdots \\ P_{L_I} \end{bmatrix} \begin{bmatrix} \boldsymbol{v}^1 & \cdots & \boldsymbol{v}^{M,} \end{bmatrix} \quad (4)$$

where U is the $IN_{dim} \times M$ orthogonal matrix, V is the $M \times M$ orthogonal matrix, S is the $M \times M$ diagonal matrix composed of singular values, P_{L_i} is the $N_{dim} \times M$ submatrix of US, and $\boldsymbol{v}^m$ is the M dimensional column vector.

The vector $\boldsymbol{v}^m$ is an intrinsic feature vector of the mth subject and is common to all strides. The submatrix P_{L_i} is
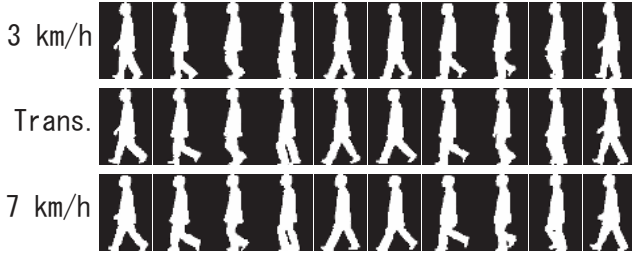
Figure 3. Silhouette transformation result from 3 km/s to 7 km/s

Figure 4. Silhouette transformation result from 6 km/s to 2 km/s

a projection matrix from the intrinsic vector $\boldsymbol{v}$ to the feature vector for stride L_i, and is common to all subjects. Thus, the feature vector $\boldsymbol{v}_{L_i}^m$ for the stride L_i of the mth subject is represented as

$$\boldsymbol{a}_{L_i}^m = P_{L_i}\boldsymbol{v}^m. \tag{5}$$

The intrinsic feature vector of the mth subject is estimated by applying the least square method to eq. (5) as

$$\hat{\boldsymbol{v}}^m = P_{L_i}^+ \boldsymbol{a}_{L_i}^m, \tag{6}$$

where $P_{L_i}^+$ is pseudo inverse matrix of P_{L_i}. Thus, feature vector transformation from a reference stride L_{ref} to a target stride L_{trg} is now easily obtained as

$$\hat{\boldsymbol{a}}_{L_{trg}}^m = P_{L_{trg}} P_{L_{ref}}^+ \boldsymbol{a}_{L_{ref}}^m, \tag{7}$$

Note that this transformation can be applied to an identification target subject as well as the training subjects because the transformation matrix $P_{L_{trg}} P_{L_{ref}}^+$ does not include terms related to subjects.

3.4. GSV reconstruction

The GSV is reconstructed by combining the unchanged static features and the transformed dynamic features. First, the upper body silhouette region is copied as it is. As for the lower body, each model part region is translated and rotated by considering joint angle changes by transformation and silhouette region inside the model region is copied after the same translation and rotation. The results of the GSV reconstruction for widening and narrowing stride are shown in Figs. 3 and 4 respectively.

4. Experiments

4.1. Method

We made experiments of gait identification under speed variations. In this experiment, the appearance-based frequency-domain feature[9] is adopted as gait features for person identification because of its efficiency. An overview of feature extraction and matching processes is briefly described as follows.

First, one dimensional Fourier analysis is applied to a time series of silhouette signal at each pixel in the GSV and an amplitude spectra is subsequently calculated for 0-, 1-, and 2-times frequencies. We can reconstruct an image for each frequency component for visibility as shown in Fig. 5, and this is treated as a unit of frequency-domain feature. Then, Euclidean distance between frequency-domain features of a gallery and a probe is computed as a matching measure. If a sequence contains multiple gait periods and it provides multiple features, we integrate the distances for all the combinations of the multiple features by minimum or median to provide a final matching score.

The two types of experimental datasets (call them dataset 1 and dataset 2 respectively) were arranged to see the effect of the number of subjects for training the transformation model. The number of training subjects in dataset 1 and 2 are 14 and 9 respectively, and the both datasets have 20 testing subjects in common. The speed variation in the database is range from 2 km/h to 7 km/h at 1 km/h interval.

As a performance evaluation measure, we adopted Equal Error Rate (EER) which is defined in the Receiver Operating Characteristic (ROC) curve[12]. The ROC curve shows a relation between False Rejection Rate (FRR) and False Acceptance Rate (FAR) when an acceptance threshold changes in the context of a verification scenario. Naturally, the lower FAR, FRR, and EER mean the higher performance.

4.2. Results

First, the matching results of the frequency-domain features before and after transformation are shown in Fig. 5. We can see that the differences of the frequency-domain features are reduced by the stride transformation in both cases.

Then, the overall performance is shown by EERs with and without transformation in Fig. 6 for each probe. In these figures, the horizontal and vertical axes mean gallery speed and EER. The averaged EERs on different speed combinations were reduced from 15.0% without transformation to 10.9% (4.1% improvement) and 11.0% (4.0% improvement) with transformation in dataset 1 and 2 respectively. As a whole, the improvement becomes larger as the speed difference between the gallery and the probe becomes larger, and in particular the largest EER improvements for

Trans.　　　　4 km/h

6 km/h

Frq.　0　1　2　0　1　2

(a) Transformation from 4 km/h (gallery) to 6 km/h (probe)

Trans.　　　　6 km/h

2 km/h

Frq.　0　1　2　0　1　2

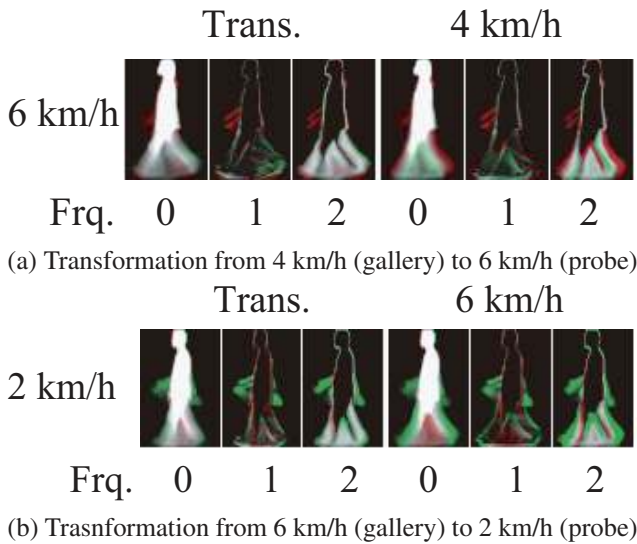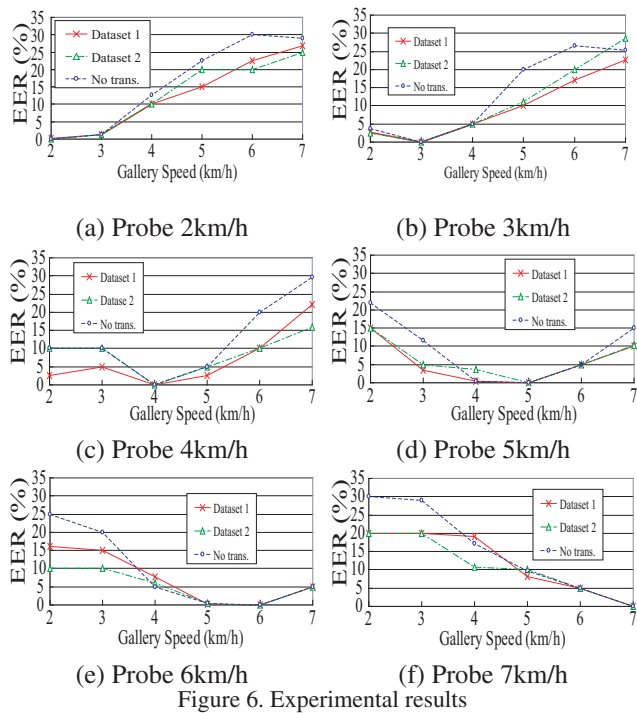(b) Trasnformation from 6 km/h (gallery) to 2 km/h (probe)

Figure 5. Reduction of feature difference by stride transformation. Red: exist only in gallery/transformed features, green: exist only in a probe feature, white: exist in both features. Less red and green regions mean better matching.

(a) Probe 2km/h　　　　(b) Probe 3km/h

(c) Probe 4km/h　　　　(d) Probe 5km/h

(e) Probe 6km/h　　　　(f) Probe 7km/h

Figure 6. Experimental results

a specific pair of gallery and probe speed in dataset 1 and 2 are 10.0% and 15.0% respectively.

4.3. Comparison with related works

In this section, the proposed method is compared with the related works by Tanawongsuwan et al.[14] and Liu et
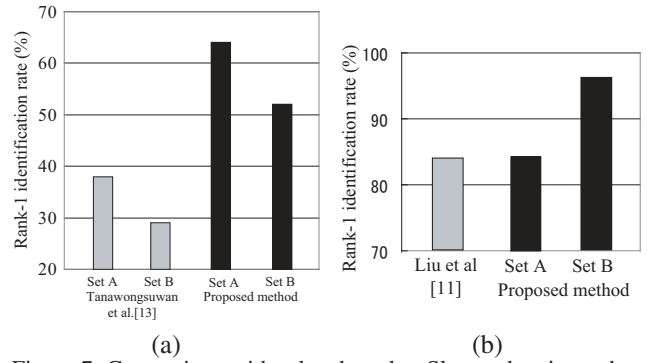

(a)　　　　　　　(b)

Figure 7. Comparison with related works. Slower data is used as a gallery in Set A, and vice versa in Set B. (a) Tanawongsuwan et al.[14] match between 2.5 km/h and 5.8 km/h, while the proposed method match between 2 km/h and 6 km/h. (b) Liu et al.[8] match between 3.3 km/h and 4.5 km/h, while the proposed method match 3 km/h and 4 km/h.

al.[8]. Because the rank 1 identification rate in the Cumulative Match Characteristic (CMC) curve was used in the related works, we also used the same performance measure in the comparison experiments.

When the CMC curve is used for performance evaluation, gallery size makes a deep impact on performance, in more detail, the larger gallery size means more difficult identification problem setting. Thus, almost the same-size gallery should be used among the methods in common for fair comparison. The gallery sizes of the related works[14][8] are 24 and 25 respectively, and therefore 25 subjects are assigned for testing while the remaining 9 subjects are used for training the transformation model in the proposed method. Comparison results are shown in Fig. 7.

Speed variations in [14] are 2.5 km/h, 3.6 km/h, 4.7 km/h, and 5.8 km/h. Focused on the largest speed difference combination (2.5 km/h and 5.8 km/h), the rank 1 identification rates are approximately 40% for 2.5 km/h gallery and 30% for 5.8 km/h gallery respectively. On the other hand, in case of the proposed method, for larger speed difference of 2 km/h and 6 km/h, the rank 1 identification rate is 64% for 2 km/h gallery and 52% for 6 km/h gallery respectively. This is because we use all of the frames for feature extraction for matching while only five key frames (three double support frames and two single support frames) are used for matching in [14].

Speed variations in [8] are 3.3 km/h and 4.5 km/h. Since the speed difference is almost 1 km/h, we compare with a combination of 3 km/h and 4 km/h in the proposed method. While the rank 1 identification rate in [8] is approximately 84%, that in the proposed method is 84% for 3 km/h gallery and 96% for 4 km/h gallery. This is because we consider both time normalization and stride transformation, while only time normalization is considered in [8].

5. Conclusion

We propose a method of gait silhouette transformation from one speed to another to cope with walking speed changes in gait identification. Because dynamic features are dependent on speed changes while static features are independent of speed changes, the dynamic and static features are separately extracted by human model fitting to the silhouette sequence. Next the dynamic features are transformed to the target speed by a factorization-based transformation model and then the GSV for the target speed is reconstructed by combining the unchanged static features and the transformed dynamic features. Finally, for the performance evaluation with and without transformation, gait identification using frequency-domain features were made for speed variation gait dataset ranging from 2 km/h to 7 km/h at 1 km/h interval. As a result, averaged EER for different speed combinations improved from 15.0% to 10.9%.

A future direction is extension of the proposed method to various gait styles such as running by introducing full body model fitting, gait style classification, and gait style-specific stride transformation models. Our faster gait dataset ranging from 8 km/h to 10 km/h at 1 km/h interval can be used for this purpose.

References

[1] C. BenAbdelkader, R. Cutler, and L. Davis. Stride and cadence as a biometric in automatic person identification and verification. In *Proc. of IEEE International Conference on Automatic Face and Gesture Recognition*, page 372, Washington, D.C., 2002. IEEE Computer Society. 1

[2] S. R. Das, R. C. Wilson, M. T. Lazarewicz, and L. H. Finkel. Gait recognition by two-stage principal component analysis. In *Proc. of International Conference on Automatic Face and Gesture Recognition*, pages 579–584, Washington, DC, USA, 2006. IEEE Computer Society. 1

[3] D. M. Gavrila. The visual analysis of human movement: A survey. *Computer Vision and Image Understandings*, 73(1):82–98, 1999. 2

[4] R. Gross and J. Shi. The cmu motion of body (mobo) database. Technical Report CMU-RI-TR-01-18, Robotics Institute, Pittsburgh, PA, Jun. 2001. 1, 2

[5] M. A. Hossain, Y. Makihara, J. Wang, and Y. Yagi. Clothes-invariant gait identification using part-based adaptive weight control. In *Proc. of the 19th Int. Conf. on Pattern Recognition*, Tampa, Florida USA, Dec. 2008. 1

[6] Y. Liu, R. Collins, and Y. Tsin. Gait sequence analysis using frieze patterns. Technical Report CMU-RI-TR-01-38, Robotics Institute, Pittsburgh, PA, Dec. 2001. 1, 2

[7] Z. Liu and S. Sarkar. Simplest representation yet for gait recognition: Averaged silhouette. *Pattern Recognition, International Conference on*, 4:211–214, 2004. 1

[8] Z. Liu and S. Sarkar. Improved gait recognition by gait dynamics normalization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(6):863–876, 2006. 1, 2, 5

[9] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, and Y. Yagi. Gait recognition using a view transformation model in the frequency domain. In *Proc. of the 9th European Conf. on Computer Vision*, volume 3, pages 151–163, Graz, Austria, May 2006. 1, 2, 3, 4

[10] Y. Makihara and Y. Yagi. Silhouette extraction based on iterative spatio-temporal local color transformation and graph-cut segmentation. In *Proc. of the 19th Int. Conf. on Pattern Recognition*, Tampa, Florida USA, Dec. 2008. 2

[11] S. D. Mowbray and M. S. Nixon. Proc. of international conference on automatic gait recognition via fourier descriptors of deformable objects. In *Audio Visual Biometric Person Authentication*, pages 566–573. Springer, 2003. 1

[12] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The feret evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, 2000. 4

[13] S. Sarkar, J. Phillips, Z. Liu, I. Vega, P. Grother, and K. Bowyer. The humanid gait challenge problem: Data sets, performance, and analysis. *Trans. of Pattern Analysis and Machine Intelligence*, 27(2):162–177, 2005. 1, 2

[14] R. Tanawongsuwan and A. Bobick. Modelling the effects of walking speed on appearance-based gait recognition. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 2:783–790, 2004. 1, 2, 5

[15] D. K. Wagg and M. S. Nixon. On automated model-based extraction and analysis of gait. In *Proc. of International Conference on Automatic Face and Gesture Recognition*, pages 11–16. IEEE Computer Society, 2004. 1

[16] L. Wang, H. Ning, T. Tan, and W. Hu. Fusion of static and dynamic body biometrics for gait recognition. *Computer Vision, IEEE International Conference on*, 2:1449, 2003. 1

[17] L. Wang, T. Tan, H. Ning, and W. Hu. Silhouette analysis-based gait recognition for human identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1505–1518, 2003. 1

[18] The ou-isir gait database, Mar. 2010. http://www.am.sanken.osaka-u.ac.jp/GaitDB/index.html. 2

[19] C. Yam, M. S. Nixon, and J. N. Carter. Automated person recognition by walking and running via model-based approaches. *Pattern Recognition*, 37(5):1057–1072, 2004. 1

[20] R. Zhang, C. Vogler, and D. Metaxas. Human gait recognition. *Computer Vision and Pattern Recognition Workshop*, 1:18, 2004. 1