IEICE TRANS. INF. & SYST., VOL.E95–D, NO.2 FEBRUARY 2012

PAPER

Robust Gait-Based Person Identification against Walking Speed Variations

Muhammad Rasyid AQMAR[†a)], ***Nonmember***, **Koichi SHINODA**[†b)], ***Senior Member***, ***and*** **Sadaoki FURUI**[†c)], ***Fellow***

SUMMARY Variations in walking speed have a strong impact on gait-based person identification. We propose a method that is robust against walking-speed variations. It is based on a combination of cubic higher-order local auto-correlation (CHLAC), gait silhouette-based principal component analysis (GSP), and a statistical framework using hidden Markov models (HMMs). The CHLAC features capture the within-phase spatio-temporal characteristics of each individual, the GSP features retain more shape/phase information for better gait sequence alignment, and the HMMs classify the ID of each gait even when walking speed changes nonlinearly. We compared the performance of our method with other conventional methods using five different databases, SOTON, USF-NIST, CMU-MoBo, TokyoTech A and TokyoTech B. The proposed method was equal to or better than the others when the speed did not change greatly, and it was significantly better when the speed varied across and within a gait sequence.

key words: *gait recognition, CHLAC features, GSP features, hidden Markov models*

1. Introduction

Human gait refers to the motion of an individual characterized by his/her spatio-temporal movement while walking. The study of gait analysis in relation to human identification has gained momentum in recent years. Several studies in the field of psychophysics have indicated that humans are capable of recognizing a person's characteristics, such as IDs and genders, from only limited information on gait cues that have been visualized by means of small light bulbs attached to body joints [1], [2].

Automatic person identification using human gait has been extensively studied [3]–[5]. Philips *et al.* [3] used the difference in binary silhouettes between frames. Huang *et al.* [4] used optical flow as features and derived Eigengait. Kale *et al.* [5] used exemplar-based silhouettes and hidden Markov models (HMMs) to represent structural and dynamic characteristics of gait sequences.

People can walk at various speeds in real life, and their motion changes *nonlinearly* in terms of its speed [6]. Speed variations can appear across and within a gait sequence, and they significantly affect the performance of gait-based person identification. Therefore, many studies have been conducted to build a gait-based person identification system robust against variations in walking speed [7]–[9]. Unfortunately, most of those studies assume that the speed change is linear across the sequence, and hence fail to address the problem of nonlinear speed changes.

In this paper, we propose a novel method of identifying humans from their gait under speed variations, where we combine cubic higher-order local auto-correlation (CHLAC) features and a statistical HMM framework [10]. We also employ the concatenation of CHLAC and gait silhouette-based principal component analysis (Gait-Silhouette-PCA, or GSP) as features and combine them with HMM. CHLAC captures shape and motion characteristics for discriminating the subjects accurately. GSP retains more shape information of the subjects to distinguish different gait phases more precisely. HMM is able to match sequences that have different speeds. We expect that this combination can perform better than using them separately.

This paper is organized as follows. Section 2 presents previous work on gait recognition related to the problem with speed variations. Section 3 reviews CHLAC-based features, Sect. 4 explains our method combining CHLAC features, GSP features and HMMs. Section 5 reports the results obtained from our experiments on the proposed method and Sect. 6 concludes the paper.

2. Related Work

The problem with speed variations in gait can be further divided into two sub-problems:

1. Finding a feature that is invariant against speed variations,
2. Preventing misalignment between time-sequence patterns.

The CMU MoBo database [7], which consists of gait data with different speeds, was built to tackle these problems. Several studies have been done using this database. For example, Zhao *et al.* [8] proposed fractal-scale wavelet moments. Lee *et al.* [9] proposed a shape variation-based frieze pattern.

Kobayashi *et al.* [11] proposed three-way (x-, y-, and time-axis) autocorrelation features that effectively represent spatio-temporal local geometric characteristics of human motions. It is called cubic higher-order local auto-correlation (CHLAC). When the shape of a human is not changed significantly, CHLAC is expected to be robust

Manuscript received July 25, 2011.
Manuscript revised October 25, 2011.

[†]The authors are with the Tokyo Institute of Technology, Tokyo, 152–8552 Japan.
a) E-mail: rasyid@ks.cs.titech.ac.jp
b) E-mail: shinoda@cs.titech.ac.jp
c) E-mail: furui@cs.titech.ac.jp

DOI: 10.1587/transinf.E95.D.668

against variations in walking speed. This is because it only uses the sum of local features over a gait sequence, and thus, does not explicitly use the phase information of the gait. To the best of our knowledge, this method outperforms all other gait-recognition methods.

Most of these methods including CHLAC, however, have focused on the first sub-problem and not dealt with the second sub-problem. Their performance may degrade greatly because of misaligned gait cycles and/or phases when walking speed varies significantly.

It is well known that human motion changes nonlinearly in terms of its speed [6]. A straightforward way of tackling the second sub-problem is to find a way of estimating the nonlinear time-warping function between two time-sequence patterns with different speeds. Veeraraghavan *et al.* [12] used Dynamic time warping (DTW), which is a well-known method of explicitly estimating the warping function. A hidden Markov model (HMM) is an extension of DTW to a probabilistic framework. HMM-based gait recognition has often been studied [5], [13] to tackle this second sub-problem. The mixture of distributions (Gaussian mixture) is often used as an output probability in HMMs in order to achieve robustness against variations in observation features. However, few studies based on DTW and HMMs have directly focused on the first problem (to utilize robust features) of speed variations.

Some other studies focused on modeling shape variation. Tanawongsuwan *et al.* [14] proposed a method based on the silhouette transformation by normalizing the stride length of a double-support pose and keyframes similarity. Tsuji *et al.* [15] improved the transformation not only by using the stride normalization method but also by adding the time synchronization. Since we mainly focus on the speed variation itself, we will leave the shape variation problem for our future research.

3. CHLAC Features

CHLAC features are shape and motion features extracted from local autocorrelation [11]. One of their most important properties is their shift invariance. They do not change if the position of a person varies inside a frame image.

Let $f(x, y, t)$ represent pixel intensity on the image region, where x and y are pixel coordinates in one frame image, and t is the time index. Each of the N-th order autocorrelation functions is defined as:

$$R_N(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N) = \sum_{x,y,t \in D_s} f(x, y, t) f(x + a_{1x}, y + a_{1y}, t + a_{1t}) \ldots f(x + a_{Nx}, y + a_{Ny}, t + a_{Nt}), \quad (1)$$

where $\boldsymbol{a}_i$ $(i = 1, \ldots, N)$ is a displacement vector from the reference point, $\boldsymbol{r} = (x, y, t)$. A set $(\boldsymbol{r}, \boldsymbol{r} + \boldsymbol{a}_1, \ldots, \boldsymbol{r} + \boldsymbol{a}_N)$ represents a local mask pattern. Figure 1 shows their examples. D_s is a spatio-temporal region where the correlation coefficient for each pixel are summed up. The size of D_s is $m \times n \times T$, where m and n is the height and the width of the region, and T is the time window width to be optimized in the experiments. For each i, $\boldsymbol{a}_i = (a_{ix}, a_{iy}, a_{it})$. a_{ix} is $\pm\Delta x$ or 0, a_{iy} is $\pm\Delta y$ or 0, and a_{it} is $\pm\Delta t$ or 0, where Δx and Δy denote the spatial displacement in pixels and Δt denotes the frame interval in frames. Here, they use the same value for Δx and Δy and denote this as Δr. When the order of correlation is $N = 0$, $N = 1$, and $N = 2$, the numbers of mask patterns (the dimensions of a CHLAC feature vector) are 1, 14, and 251 respectively. The spatial displacement is set to $\Delta r \leq 16$ which is corresponding to the upper bound width of human body in an image. They assume the relation $\Delta r = 2\Delta t$ by observing the walking trajectories of lower portion (below the torso) movement of the subjects. The trajectories of walkers on the x-axis are plotted against the time-axis. For each stride, each left and right leg movement forms a trajectory with a particular gradient ($XT-$ slice). Each walker has his/her own trajectory's gradient represented by $(\Delta t/\Delta r)$ parameters. In this sense, the parameters represents the rate or speed of the walker. Let the parameter pair, $(\Delta t, \Delta r) = (k, 2k)$, be denoted as u_k, where $k = 1, \ldots, K$. They prepare several pairs with different k.

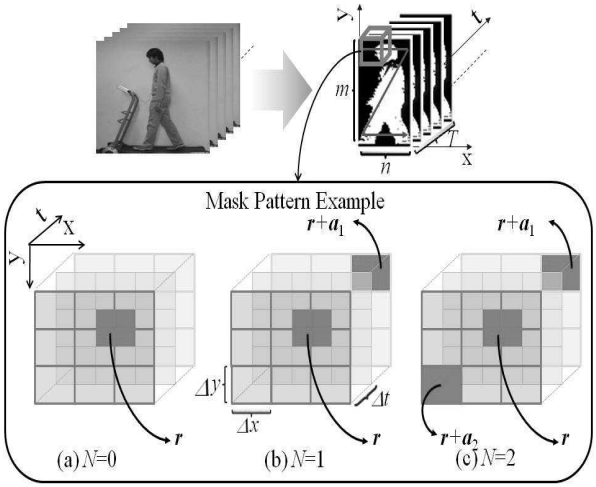


Fig. 1 Examples of mask pattern in CHLAC.

Then, the CHLAC features are mapped to the $(c-1)$-dimensional (c is the total number of classes) feature vector using Fisher discriminant analysis (FDA) to better separate classes in the feature space. We call the resulting features CHLAC+FDA.

In the original framework [11], a k-nearest neighbor (k-NN) classifier was used where the Euclidean distance was used as the distance measure between feature vectors. The number of neighbors k was set to 10. The window width, T, was set to 30.

Next, the number of nearest neighbors belonging to each ID i, $M_k(i)$, is counted for each of parameter pairs u_k. Then, the ID i with the maximum number of neighbors over all parameter pairs is selected:

$$\hat{i} = \arg\max_i M_k(i). \quad (2)$$

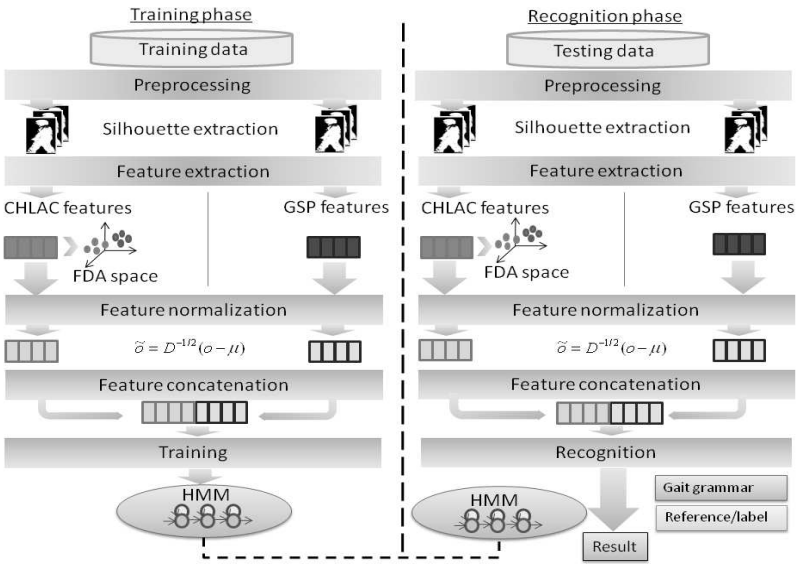


Fig. 2 Gait model training and continuous gait recognition.

The same process is repeated for all the frames. Finally, the ID i that most frequently appears over all the frames is selected.

Since all the possible parameter pairs related to walking speed (u_k) are used during the recognition process, CHLAC+FDA features are relatively robust against variations in speed across samples when changes in speed are almost constant in each gait sequence. When walking speed varies within one gait sequence, however, optimal parameter pair may change for some cycles and/or phases. This may degrade recognition performance.

4. Combination of CHLAC+FDA and GSP with HMMs

We use CHLAC+FDA and Gait-Silhouette-PCA (GSP) concatenation as observation vectors in our scheme and use an HMM as a classifier instead of the k-NN classifier in [11]. HMMs have often been used in speech recognition [16] since they are powerful for modeling time-varying sequences of patterns. CHLAC+FDA discriminates accurately between classes, GSP distinguishes gait phases more precisely, while HMMs have excellent properties to match sequences that have different speeds. We expect their combination will be robust against speed variations, even when the speed varies within a gait sequence. Figure 2 illustrates our framework.

4.1 Feature Extraction

We extract silhouette images from the video frames in the preprocessing stage for feature extraction. Since CHLAC+FDA features do not have much gait phase/cycle information, it is difficult to train HMMs solely with CHLAC+FDA features. Therefore, we not only extract CHLAC+FDA features from silhouette images, but also another feature which have more explicit phase information, such as silhouette features by principal component analysis (PCA). We apply PCA to a set of vectors of all binary silhouette images in training data and calculate the eigenvectors. We call the resulting features GSP.

Kobayashi *et al.* [11] set the window length, T, for CHLAC+FDA at 30 (frames), which roughly corresponded to the duration of a complete gait cycle. In our approach using HMMs, this window length should be smaller to capture features at a certain phase in a gait cycle. Too small window length (e.g. T=1) is, however, not enough to capture phase information. We set T at five (frames) according to the results from our preliminary experiments which will be shown in Sect. 5.2.

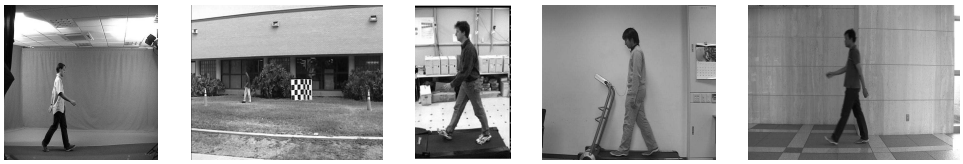
4.2 Feature Normalization and Concatenation

We concatenate CHLAC+FDA features and GSP features to make an input feature vector for HMMs (feature level fusion). Then, the input vector o is normalized as follows:

$$\tilde{o} = D^{-\frac{1}{2}}(o-\mu), \tag{3}$$

where μ and D are the mean vector and the covariance matrix respectively calculated for each gait sequence. We ignore off-diagonal elements. The purpose of normalization is not only to make the value of feature components lie within similar dynamic ranges but also for suppressing the outlier or noise in the observation features.

In [11], they used several parameter pairs $(\Delta t, \Delta r) = (k, 2k)$ where $k = 1, \ldots, K$. In our method, we only use one parameter pair $(\Delta t, \Delta r)$=(2,4) which performed the best among those pairs in our preliminary experiment (see Sect. 5.2 for detail).



(a) SOTON (b) USF-NIST (c) CMU-MoBo (d) TokyoTech A (e) TokyoTech B

Fig. 3 Examples of samples in five gait databases.

4.3 Recognition

We prepare one HMM for a half-gait cycle assuming there is symmetry between the first and the second half of the cycle in the sagittal plane view. Its topology is left-to-right without any skips. We use a mixture of Gaussian distributions as an output probability. We set the number of states in an HMM to eight. This number gave the best performance in our preliminary experiment.

We employ a *continuous* gait-recognition framework which we allow the transition from the exit state to the entry state. A task grammar is represented in an extended Backus-Naur form notation as $(< g_1 > | < g_2 > | \ldots | < g_n >)$.

Given a gait sequence of observation vectors, $O = o_1 \ldots, o_f$ (f is the number of frames), the probability of gait ID g_i is:

$$P(g_i|O) = \frac{P(O|g_i)P(g_i)}{P(O)}, \tag{4}$$

where $P(g_i)$ is the prior of gait ID g_i, which is assumed to be uniform over all IDs. The most probable gait ID $\hat{i}$ can be selected as

$$\hat{i} = \arg\max_i P(O|g_i). \tag{5}$$

We simultaneously obtain the probability, $P(O|g_i)$, and gait-cycle segmentation by using the Viterbi algorithm.

5. Experiments

We first compared our method with other conventional approaches under conditions where walking speed did not change significantly. Second, we evaluated it under conditions where the walking speed was changed.

5.1 Database

For the first evaluation, we used the University of Southampton's (SOTON) large database [17] (115 subjects) and the University of South Florida's (USF)-NIST database Probe A [3] (71 subjects). Walking speed in both of them were the same.

For the second evaluation, we used the Carnegie Mellon University-Motion of Body (CMU-MoBo) [7] (25 subjects), our own TokyoTech database A (30 subjects), and TokyoTech database B (15 subjects). These databases include subjects walking at various speeds.

Table 1 Average (μ) and standard deviation (σ) of half-gait cycle periods (sec) in SOTON and USF-NIST Probe A databases.

	SOTON		USF-NIST	
	μ	σ	μ	σ
Training set	0.55	0.07	0.59	0.08
Testing set	0.55	0.06	0.58	0.08

The SOTON large database [17] consists of 115 subjects. It has one variation, in the camera view (left and right). This database was collected indoors with a uniform background, as can be seen in Fig. 3 (a). The speed variations were very small. We measured the average period of a half gait cycle for all subjects in the database (Table 1), and observed that the speed variations in this database were not significant. The training set for the SOTON database was the data recorded by the left-view camera, while the testing set was recorded by the right-view camera.

The USF-NIST database [3] was collected in an outdoor environment with a complex background, as can be seen in Fig. 3 (b). The number of subjects was 71. The database has three types of variations: the surface type, camera view, and shoes. The database was constructed based on a combination of these variations. The speed variations in this database were also not very significant (Table 1). In our experiments, we used as the training set those data recorded from the grass surface, using shoe-type A, and from the right-camera view (Gallery). As our test set, we selected data recorded from the left-camera view, that recorded on the grass surface, in which the subjects wore shoe-type A (Probe A). The difference of the training and testing set was only the camera-view.

The CMU-MoBo database [7] consists of 25 subjects. The database has six types of camera views: $0°, 45°, 90°$, $180°, 225°$, and $315°$, and four types of walking conditions: slow, fast, incline walking, and walking with a ball recorded on a treadmill. One data sample can be seen in Fig. 3 (c). We only used the slow and fast speed with the sagittal plane view ($90°$). The slow speed was recorded on the treadmill at 3.3 km/hr, and the fast speed was recorded on a treadmill at 4.5 km/hr. There was only one sample for each speed for each subject. The experiment using this database was conducted with four types: (S-S) the slow sequence was divided into 50% for the training set and 50% for the testing set, (F-F) the fast speed sequence was divided into 50% for the training set and 50% for the testing set, (S-F) the slow speed was used as the training set

Table 2 TokyoTech database A.

Speed type	Slow	Normal	Fast	Mixed
Speed (km/hr)	2	3	4.5	3 and 4.5
No. of samples	605	550	447	300

while the fast speed was used as the testing set for a whole sequence, (F-S) the fast speed was used as the training set while the slow speed was used as the testing set for a whole sequence.

The TokyoTech database A was constructed by ourselves that included 30 subjects walking at various fixed speeds. The gait data were categorized into four types: slow, normal, fast, and mixed (Table 2). A treadmill was used to ensure that all subjects walked at exactly the same speed in each speed category. The sagittal plane of the subjects was taken using a 30-fps video camera with a pixel-frame size of 480×720. The setting for recording can be seen in Fig. 3 (d). To eliminate the possible effect of shoe differences, all subjects wore shoes with the same shape and color. The total length of the video data recorded for the 30 subjects was around 324 min.

We divided the slow data into two parts, 60% for training (363 samples) and 40% for testing (242 samples). The training set only consisted of slow data. The testing set consisted of the rest of the slow data and data with the other three speeds. While the testing set in the CMU-MoBo database only included one sequence for one subject for each type of speed, TokyoTech database A provided several numbers of sequences so that the results from the evaluation become more statistically convincing. In each evaluation, one sample was a gait sequence that contained five gait cycles from one subject.

We also manually created mixed data, where there were two different speeds within one gait sequence. The purpose was to evaluate what effect variations had within one gait sequence. We concatenated two different speeds in a gait sequence in the point when both left and right leg passes each other from the sagittal plane point of view. 150 samples are constructed by concatenating three gait cycles from the normal data and two gait cycles from fast data, and 150 samples are constructed from the concatenation of three gait cycles from the fast data and two gait cycles from the normal data.

The TokyoTech database B. To evaluate the methods in a more realistic environment, we constructed TokyoTech database B that included 15 subjects walking at slow and fast speed. The subject walked on the ground floor with a non-uniform background and illumination condition.

The sagittal plane of the subjects was taken using video camera with a pixel-frame size of 720×576 and 30 fps frame rate. We categorized the data into four different walking style (Table 3): (A) the subjects walked at slow speed, (B) the subjects walked at faster speed, (C) the subjects at first walked at slow speed in the first half of the sequence, and then walked faster in the next half of the sequence, (D) the subjects at first walked at fast speed in the first half of sequence, and then walked slower in the next half of the sequence. The number of samples in (A), (B), (C), (D) are 300, 300, 150, 150 samples respectively.

Table 3 TokyoTech database B.

Speed type	Slow	Fast	SlowFast	FastSlow
No. of samples	300	300	150	150

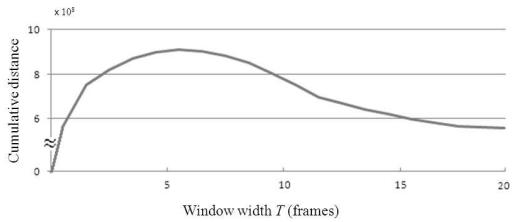


Fig. 4 Cumulative distance between neighbouring CHLAC features. For each T, we summed up the Euclidean distance of two successive CHLAC feature vectors over all the training samples.

We divided Data (A) into two parts, and used 75% (225 samples) for training and 25% (75 samples) for testing. The training set only consisted of slow data. The testing set consisted of (B), (C), (D), and the rest of (A).

5.2 Experimental Setup

We assumed that a background image was available in the preprocessing stage. After a certain threshold for the intensity of each pixel was set, the foreground pixels were extraced as a binary silhouette image. Then, the bounding box around the silhouette was resized into $m \times n$ pixels. Silhouette images were kept in the center region (registered). We set $m = 128$ and $n = 88$ following the case in NIST dataset.

We used the 0^{th} to 2^{nd} order CHLAC features and applied FDA. FDA is carried out using all training data[†]. The window width, T, of CHLAC features in our approach should be wide enough to capture a certain phase in a gait cycle and should maintain a steady pose during that period. We carried out a preliminary experiment to select T using cumulative distances between neighbouring CHLAC features (Fig. 4.). We expected that the gait-phase information between two successive CHLAC feature vectors could be more easily distinguished if the distance between them was larger. We summed up the Euclidean distances of two successive CHLAC feature vectors for each T over all the training samples. $T = 5$ (five) frames gave the highest cumulative distance. We used $T = 5$ in the following experiments.

Since each walker has his/her own trajectory's gradient represented by u_k parameters, we measured the average absolute gradient of the trajectories for each subject in the XT-

[†]We apply the FDA result for all gait *phases*. This is because we have to use the same features for all phases in our continuous recognition framework using HMMs.

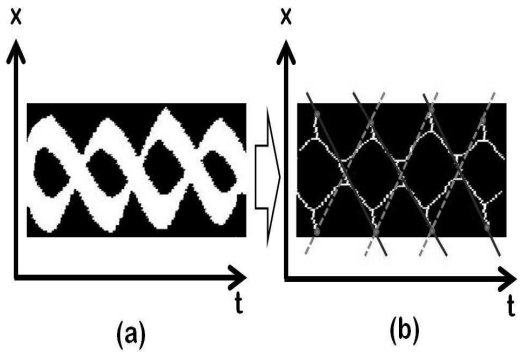


Fig. 5 The XT-slice plot from silhouette (a) and after skeletonization (b). The gait gradients are depicted by solid and dashed lines in (b).

slice (Fig. 5). We used the same constraint as in [11]. We set K to 8 in our experiment for parameter pairs. Thus, parameter pairs were set to be (1,2), (2,4), (3,6), (4,8), (5,10), (6,12), (7,14), and (8,16). We tested those parameter pairs and selected parameter pair (2,4) since it performed best in our experiment.

The dimension of CHLAC+FDA feature is 114 for SOTON, 70 for USF-NIST, 24 for CMU-MoBo, 29 for the TokyoTech A, and 14 for the TokyoTech B. These dimensions were automatically determined by the number of subjects to be classified. We set the dimension of GSP features to be $\frac{1}{2}(c-1)$ where c is the number of classes/subjects for each database. We found that if the GSP dimension was larger than CHLAC+FDA dimension, the performance decreased. This might be because the subject discrimination information from CHLAC+FDA became less dominant than GSP. We determined the GSP dimension in feature concatenation based on grid search among $c-1$, $\frac{3}{4}(c-1)$, $\frac{1}{2}(c-1)$, and $\frac{1}{4}(c-1)$ dimensions using two-fold cross-validation. We selected $\frac{1}{2}(c-1)$ dimension which gives the best performance in average. We used 57-dimensional GSP features for SOTON, 35 for USF-NIST, 12 for CMU-MoBo, 15 for the TokyoTech database A, and 7 for the TokyoTech database B. The cumulative contribution rates of the eigenvectors for the concatenated GSP dimension were 86.2%, 81.1%, 67.0%, 77.4%, and 64.3% for SOTON, USF-NIST, CMU-MoBo, TokyoTech database A, and TokyoTech database B respectively. We also used 60-dimensional GSP features, which are not combined with CHLAC+FDA features, in our evaluation using the TokyoTech database A. The cumulative contribution rate of the first 60 eigenvectors (principal components) was 86.7%.

We used a single Gaussian distribution in each state of the HMMs in our evaluation using the SOTON, USF-NIST, and CMU-MoBo due the limited number of training samples. We used a Gaussian-mixture distribution with 16 mixtures in our evaluation using the TokyoTech A and B since the number of training samples on both databases was relatively large. The number of Gaussian-mixture used for each database was determined by employing leave-one-out cross validation using the training data. We used HTK [20] to implement HMMs.

Table 4 Gait recognition accuracy (%) for SOTON and USF-NIST Probe A databases. In CHLAC+FDA-k-NN, CHLAC+FDA was used as features and k-NN as a classifier. In CHLAC+FDA+GSP-HMM, CHLAC+FDA+GSP was used as features and HMM as a classifier (proposed method).

	SOTON	**USF-NIST**
Foster *et al.* [18]	75.0	–
Tolliver *et al.* [19]	–	82.0
Kale *et al.* [5]	–	99.0
CHLAC+FDA-k-NN	98.3	100.0
CHLAC+FDA-HMM	**98.3**	**100.0**
CHLAC+FDA+GSP-HMM	**98.3**	**100.0**

We examined the segmentation of gait half-cycles to confirm how well HMM states were aligned the gait phases/cycles. We defined a gait half-cycle segment as a sequence of frames from a single-support gait pose to the next single-support gait pose. A single-support gait pose is the pose or point when the right and the left leg/foot overlap in a gait cycle. The segmentation result for each test sequence was compared to the manual labels/groundtruth. For the groundtruth, we manually marked by hand the frames which contain a single-support gait pose as segmentation boundaries. We chose the frame where one foot was completely overlapped by the other one as the segment boundary. We defined the misalignment for each boundary in a gait sequence as the number of frames difference between the groundtruth and that obtained automatically by the HMM. We then calculated the average misalignment over all gait test sequences. The average mis-alignments of frames from all databases when we used only CHLAC+FDA features and after we employed features fusion (CHLAC+FDA+GSP) were 13.4 and 12.3 frames respectively. We confirmed that segmentation was better when we combined CHLAC+FDA and GSP features.

The computation time for silhouette and CHLAC+FDA extraction was 0.13 second for each frame, and the recognition process of a gait sequence using HMM was 0.06 second using Intel Core 2 Duo 2.4 GHz with 2 GB RAM.

5.3 Results

Table 4 lists the results in SOTON and USF-NIST. CHLAC+FDA+GSP-HMM yielded one of its best results. For the SOTON database, we have presented the results for the area-based mask pattern-k-NN proposed by Foster *et al.* [18] and CHLAC+FDA-k-NN [11] for comparison. The proposed CHLAC+FDA+GSP-HMM was better than the area-based mask pattern-k-NN and equal to CHLAC+FDA-k-NN. For the USF-NIST database, we compared our method with three methods, Shape-1-NN [19], Silhouette frame-to-exemplar-distance (FED)-HMM [5], and CHLAC+FDA-k-NN [11]. The results for these methods were taken from those published in corresponding papers. We found that our method was significantly better than Shape-1-NN and almost equal to FED-HMM, CHLAC+FDA-k-NN, and CHLAC+FDA-HMM.

Table 5 Gait recognition accuracy (%) for CMU Mobo database. In CHLAC+FDA-*k*-NN, CHLAC+FDA was used as features and *k*-NN as a classifier. In CHLAC+FDA+GSP-HMM, CHLAC+FDA+GSP was used as features and HMM as a classifier (proposed method). S-S: slow as training set and slow as testing set. F-F: fast as training set and fast as testing set. S-F: slow as training set and fast as testing set. F-S: fast as training set and slow as testing set.

	S-S	**F-F**	**S-F**	**F-S**	**Average**
Kale *et al.* [5]	72.0	68.0	32.0	56.0	57.0
Lee *et al.* [9]	100.0	100.0	82.0	80.0	90.5
Liu *et al.* [13]	-	-	84.0	-	-
CHLAC+FDA-*k*-NN	100.0	100.0	96.0	96.0	98.0
CHLAC+FDA-HMM	100.0	100.0	96.0	96.0	**98.0**
CHLAC+FDA+GSP-HMM	100.0	100.0	96.0	96.0	**98.0**

Table 6 Gait-based person identification accuracy (%) for TokyoTech A. In GSP-HMM, GSP was used as features and HMM as a classifier. In CHLAC+FDA-*k*-NN, CHLAC+FDA was used as features and *k*-NN as a classifier. In CHLAC+FDA+GSP-HMM, CHLAC+FDA+GSP was used as features and HMM as a classifier (proposed method).

Speed Type	**Slow**	**Normal**	**Fast**	**Mixed**	**Average**
GSP-*k*-NN	81.8	74.6	39.5	44.1	60.1
GSP-HMM	96.7	85.0	60.0	76.3	79.5
CHLAC+FDA-*k*-NN	100.0	95.5	91.0	92.0	95.4
CHLAC+FDA-HMM	100.0	98.2	95.3	95.6	**97.3**
CHLAC+FDA+GSP-HMM	100.0	98.2	95.8	96.7	**97.6**

Table 7 Gait-based person identification accuracy (%) for TokyoTech B.

Speed Type	**Slow**	**Fast**	**SlowFast**	**FastSlow**	**Average**
CHLAC+FDA-*k*-NN	96.0	77.3	83.7	81.5	84.6
CHLAC+FDA-HMM	98.3	85.3	84.6	82.7	**87.7**
CHLAC+FDA+GSP-HMM	98.3	85.8	85.3	84.6	**88.4**

Next, we evaluated the robustness of our proposed method against speed differences using CMU-MoBo, TokyoTech A, and TokyoTech B. Table 5 shows the results for the CMU-MoBo database. Our method, CHLAC+FDA+GSP-HMM, was better than FED-HMM and the Shape-variation based Frieze pattern, and it was equal to the CHLAC+FDA-*k*-NN and CHLAC+FDA-HMM method. It was also better than Liu *et al.* [13] which also used CMU-MoBo and slightly better than Tsuji *et al.* [15] which used their own database with similar settings as the CMU-MoBo. Tsuji *et al.* [15] used their own database where the speed is at 3 km/hr as training set against 4 km/hr as testing set as an approximation for CMU-MoBo. The identification rate of our method is 96%, while Liu *et al.* [13] reported 84% and Tsuji *et al.* [15] also reported 84%. On a reverse condition where the speed at 4.5 km/hr as training set and speed at 3.3 km/hr as testing set, the identification rate of our method is 96% and the identification rate of Tsuji *et al.* [15] was also 96%.

The recognition results for TokyoTech A and B are listed in Table 6 and Table 7 respectively. Our method yielded better results than CHLAC+FDA-*k*-NN and CHLAC+FDA-HMM. When walking speed was "Mixed", 96.7% accuracy was achieved while the accuracies obtained with the CHLAC+FDA-*k*-NN and CHLAC+FDA-HMM were 92.0% and 95.6%, respectively for TokyoTech A. When walking speed was "SlowFast" and "FastSlow", 85.3% and 84.6% accuracies were achieved by our method respectively while the accuracies obtained with the CHLAC+FDA-*k*-NN were 83.7% and 81.5%, respectively, and the accuracies obtained with the CHLAC+FDA-HMM were 84.6% and 82.7%, respectively for TokyoTech B. The average recognition results of our method when using features concatenation (CHLAC+FDA and GSP) for TokyoTech A and TokyoTech B were 0.3 and 0.7 point, respectively better than when using only CHLAC+FDA. The robustness against walking-speed variations across ("Normal" and "Fast" testing sets) and within ("Mixed", "SlowFast", and "FastSlow" testing sets) sequences was confirmed.

One possible reason our method outperformed CHLAC+FDA-*k*-NN under walking-speed variations might be that CHLAC+FDA-*k*-NN assumed a constant period of a gait cycle for each subject. Because the time width T parameter of CHLAC features in CHLAC+FDA-*k*-NN was set to be close to gait cycle periods in the training set, it was not suitable when the walking speed for each subject differed in the testing set. In addition, the *k*-NN based method did not utilize any timing information.

In TokyoTech A, five subjects out of 30 subjects had average identification error rate more than 10% (The average identification rate is 87.6%). Their stride length difference between slow and fast speed is larger than that of the rest of the subjects. For the subject with the highest error rate, the stride length difference between slow and fast speed is around 30 pixels (Fig. 6). Even though our method outper-

(a) Narrow (b) Wider

Fig. 6 Stride shape variation occured between training (a) and testing (b) dataset.

formed the others, it was still insufficient when the shape of the walker changed drastically. It is still needed to minimize the effect of shape variations to further improve the identification performance.

6. Conclusion

We proposed a gait recognition method robust against speed variations based on the combination of CHLAC+FDA, GSP features and HMM. By using SOTON, USF-NIST, CMU-MoBo, TokyoTech A, and TokyoTech B, we confirmed that the proposed method performed well for different speed rates both across and within sequences. On average, our method successfully reduced the errors from 4.6% (by CHLAC+FDA-*k*-NN) to 2.4% for TokyoTech A and from 15.4% (by CHLAC+FDA-*k*-NN) to 11.6% for TokyoTech B.

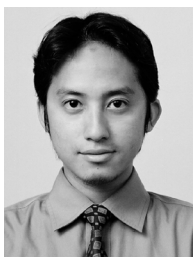
In future work, we plan to investigate ways of minimizing the influence of shape variations. Also, we would like to combine multiple CHLAC parameter pairs and apply an adaptation scheme to the HMM-based framework to further improve the recognition performance.

Acknowledgments

This work was supported by a Grant-in-Aid for Scientific Research (B) 20300063.

References

[1] G. Johansson, "Visual motion perception," Scientific American, vol.232, pp.76–88, 1975.
[2] J. Cutting and L. Kozlowski, "Recognizing friends by their walk: gait perception without familiarity cues," Bull. Psychonom. Soc., vol.9, pp.353–356, 1977.
[3] P.J. Phillips, S. Sarkar, I. Robledo, P. Grother, and K.W. Bowyer, "The gait identication challenge problem: data sets and baseline algorithm," Proc. ICPR, pp.385–388, 2002.
[4] P. Huang, C. Harris, and M. Nixon, "Human gait recognition in canonical space using spatio-temporal templates," Proc. IEEE Vision, Image Signal Processing, pp.93–100, 1999.
[5] A. Kale, A. Sundaresan, A. Rajagopalan, A. Cuntoor, N. Roy-Chowdhury, A. Kruger, and R.V. Chellappa, "Identification of humans using gait," IEEE Trans. Image Process., vol.13, no.9, pp.1163–1173, 2004.
[6] A. Bobick and R. Tanawongsuwan, "Performance analysis of time-distance gait parameters under different speeds," 4th Int. Conf. on AVBPA, pp.715–724, June 2003.
[7] R. Gross and J. Shi, "The CMU motion of body (MOBO) database," Technical Report CMU-RI-TR-01-18, Robotics Inst., Carnegie Mellon Univ., 2001.
[8] G. Zhao, L. Cui, and H. Li, "Gait recognition using fractal scale and wavelet movements," 18th IEEE ICPR, pp.453–456, 2006.
[9] S. Lee, Y. Liu, and R. Collins, "Shape variation-based frieze pattern for robust gait recognition," Proc. IEEE CVPR, pp.1–8, 2007.
[10] M.R. Aqmar, K. Shinoda, and S. Furui, "Robust gait recognition against speed variation," Proc. ICPR, pp.2190–2193, 2010.
[11] T. Kobayashi and N. Otsu, "Three-way auto-correlation approach to motion recognition," Pattern Recognit. Lett., vol.30, pp.212–221, 2009.
[12] A. Veeraraghavan, R. Chellapa, and A. Roy-Chowdury, "Rate-invariant recognition of humans and their activities," IEEE Trans. Image Process., vol.18, pp.1326–1339, 2009.
[13] Z.Y. Liu and S. Sarkar, "Improved gait recognition by gait dynamics normalization," IEEE Trans. Pattern Anal. Mach. Intell., vol.28, no.6, pp.863–876, 2006.
[14] R. Tanawongsuwan and A. Bobbick, "Modelling the effects of walking speed on appearance-based gait recognition," Proc. Comput. Vis. Pattern Recognit., pp.783–790, 2004.
[15] A. Tsuji, Y. Makihara, and Y. Yagi, "Silhouette transformation based on walking speed for gait identification," Proc. Comput. Vis. Pattern Recognit., pp.717–722, 2010.
[16] L. Rabiner and B.H. Juang, Fundamentals of Speech Recognition, Prentice Hall, 1993.
[17] J.D. Shutler, M.G. Grant, M.S. Nixon, and J.N. Carter, "On a large sequence-based human gait database," Proc. 4th International Conf. on Recent Advances in Soft Computing, pp.66–72, 2002.
[18] J.P. Foster, M.S. Nixon, and A. Prugel-Bennett, "Automatic gait recognition using area-based metrics," Pattern Recognit. Lett., vol.24, no.14, pp.2489–2497, 2003.
[19] D. Tolliver and R.T. Collins, "Gait shape estimation for identification," Proc. AVBPA, pp.734–742, 2003.
[20] S. Young, The HTK book (for HTK version 3.4), Entropic Ltd., 2009.



Muhammad Rasyid Aqmar received his B.E. in engineering physics from the Bandung Institute of Technology, Bandung, Indonesia in 2006. He received his M.E. degree in computer science from the Tokyo Institute of Technology, Tokyo, Japan, in 2009. He is currently pursuing a Ph.D. at the Tokyo Institute of Technology.

Koichi Shinoda received his B.S. in 1987, his M.S. in 1989, both in physics from the University of Tokyo, and his Dr. Eng. degree in computer science from the Tokyo Institute of Technology in 2001. In 1989, he joined NEC Corporation Japan, and was involved in research on automatic speech recognition. From 1997 to 1998 he was a visiting scholar with Bell Labs at Lucent Technologies in Murray Hill, NJ. From June to September 2001, he was a principal researcher with the Multimedia Research Laboratories of NEC Corporation. From October 2001 to March 2002, he was an associate professor at the University of Tokyo. He is currently an associate professor at the Tokyo Institute of Technology. His research interests include speech recognition, statistical pattern recognition, and human interfaces. Dr. Shinoda received the Awaya Prize from the Acoustic Society of Japan in 1997 and the Excellent Paper Award from the Institute of Electronics, Information, and Communication Engineers (IEICE) in 1998. He is an associate editor of Computer Speech and Language. He is a member of IEEE, ACM, ASJ, IPSJ, and JSAI.

Sadaoki Furui received his B.S., M.S. and Ph.D. in mathematical engineering and instrumentation physics from Tokyo University, Tokyo, Japan in 1968, 1970 and 1978. He is currently a Professor at the Tokyo Institute of Technology, Department of Computer Science. He is engaged in a wide range of research on speech analysis, speech recognition, speaker recognition, speech synthesis, and multimodal human-computer interactions and has authored or coauthored over 800 published articles. He is a Fellow of the IEEE, the International Speech Communication Association (ISCA), and the Acoustical Society of America. He has served as President of the Acoustical Society of Japan (ASJ) and the ISCA. He has served as a member of the Board of Governors of the IEEE Signal Processing (SP) Society and Editor-in-Chief of both the Transactions of the IEICE and the Journal of Speech Communication. He has received the Yonezawa Prize, Paper Award and Achievement Award from the IEICE (1975, 1988, 1993, 2003, 2003, and 2008), and the Sato Paper Award from the ASJ (1985 and 1987). He has received the Senior Award and Society Award from the IEEE SP Society (1989 and 2006), the Achievement Award from the Minister of Science and Technology and the Minister of Education, Japan (1989 and 2006), the Purple Ribbon Medal from the Japanese Emperor (2006), and the ISCA Medal for Scientific Achievement (2009). In 1993 he served as an IEEE SPS Distinguished Lecturer.