

Improved Gait Recognition by Gait Dynamics Normalization

Zongyi Liu and Sudeep Sarkar, *Senior Member, IEEE*

Abstract—Potential sources for gait biometrics can be seen to derive from two aspects: gait shape and gait dynamics. We show that improved gait recognition can be achieved after normalization of dynamics and focusing on the shape information. We normalize for gait dynamics using a generic walking model, as captured by a *population* Hidden Markov Model (pHMM) defined for a set of individuals. The states of this pHMM represent gait stances over one gait cycle and the observations are the silhouettes of the corresponding gait stances. For each sequence, we first use Viterbi decoding of the gait dynamics to arrive at one dynamics-normalized, averaged, gait cycle of fixed length. The distance between two sequences is the distance between the two corresponding dynamics-normalized gait cycles, which we quantify by the sum of the distances between the corresponding gait stances. Distances between two silhouettes from the same generic gait stance are computed in the linear discriminant analysis space so as to maximize the discrimination between persons, while minimizing the variations of the same subject under different conditions. The distance computation is constructed so that it is invariant to dilations and erosions of the silhouettes. This helps us handle variations in silhouette shape that can occur with changing imaging conditions. We present results on three different, publicly available, data sets. First, we consider the HumanID Gait Challenge data set, which is the largest gait benchmarking data set that is available (122 subjects), exercising five different factors, i.e., viewpoint, shoe, surface, carrying condition, and time. We significantly improve the performance across the hard experiments involving surface change and briefcase carrying conditions. Second, we also show improved performance on the UMD gait data set that exercises time variations for 55 subjects. Third, on the CMU Mobo data set, we show results for matching across different walking speeds. It is worth noting that there was no separate training for the UMD and CMU data sets.

Index Terms—Gait recognition, biometrics, LDA, gait shape, population HMM.

1 INTRODUCTION

COMPUTER vision-based gait biometrics has recently received a lot of attention. The genesis of the idea of being able to recognize from gait can be traced back to Cutting and Kozlowski's perception experiments [1] based on light point displays. They showed that it is possible to identify a person from the manner of walking, i.e., gait. The first effort toward *recognition from gait* in computer vision was probably done by Niyogi and Adelson in the early 1990s [2]. Over the past five years or so, significant progress has been made in terms of the diversity of gait recognition algorithms. Excellent reviews of the current state of the art can be found in [3], [4].

A meta-analysis of the identification rates reported in the recent literature can reveal outstanding hard issues in gait recognition. We consider recognition rates reported on publicly available experimental protocols and data sets (> 25 persons) such as the CMU-Mobo data set [5] (indoor, 25 subjects), UMD data set [6] (outdoor, 55 subjects), SOTON Large data set [7] (indoor and outdoor, 115 subjects), HumanID Gait Challenge data set [8] (outdoor, 122 subjects). The average identification rates for various conditions are plotted in Fig. 1. It suggests that outdoor gait recognition, recognition across walking surface-type

change, and recognition across months are all hard problems. It is reasonable to hypothesize that differing silhouette quality can possibly explain poor performance in outdoor conditions or across surface conditions, which is also accompanied by background change. However, in [9], [10], we demonstrated, based on both manual "clean" silhouettes and automatically "cleaned" silhouettes, that the poor performance cannot be explained by the silhouette quality. In this paper, we propose a new gait recognition algorithm that is able to effectively compensate for the hard covariates, such as surface, time, carrying condition, and walking speed, by normalizing the gait dynamics based on a population-based generic walking model. From the meta-analysis, we also see that performance drops with data set size, which suggests that it is imperative to demonstrate the efficacy of an idea on as large a data set as possible. Considering this, in this paper, we concentrate on recognition approaches that have been benchmarked on data sets larger than 50 subjects.

The gait of a person is a periodic activity with each gait cycle covering two strides—the left foot forward and right foot forward strides. Each stride spans the double-support stance to the legs-together stance as the legs swing past each other and back to the double-support stance. Potential sources for gait biometrics can be seen to derive from two aspects: shape and dynamics. Shape refers to the configuration or shape of the people as they perform different gait phases. Dynamics refers to the rate of transition between these phases and is usually the aspect one refers to when one talks about gait in traditional problem contexts, such as biomechanics or human motion recognition. However, we have observed that gait dynamics is vulnerable to changes in

• The authors are with the Computer Science and Engineering Department, University of South Florida, 4202 E. Fowler Ave, ENB 118, Tampa, FL 33620. E-mail: {zliu4, sarkar}@cse.usf.edu.

Manuscript received 30 Sept. 2004; revised 7 June 2005; accepted 5 Oct. 2005; published online 13 Apr. 2006.

Recommended for acceptance by J. Phillips.

For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-0519-0904.

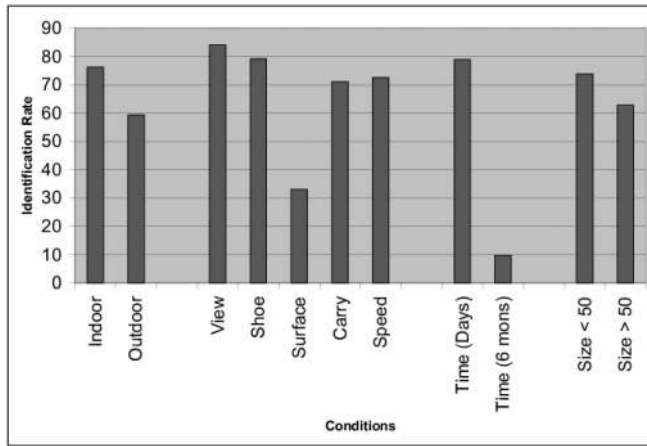


Fig. 1. Meta-analysis of gait identification rates as reported in the literature for different conditions. The average of the rates reported in [8], [11], [12], [7], [13], [6], [14], [15], [16], [17], [18], [3] are plotted for different conditions.

factors (covariates) such as walking surface, walking speed, and, to some extent, carrying conditions. Due to this potential high intrasubject variability, dynamics might not be a stable source for biometric information. Similar conclusions were drawn by Veeraraghavan et al. [19], who found that “shape is more significant for person identification than kinematics.”

Gait recognition approaches are basically of three types: 1) temporal alignment-based, 2) static parameter-based, and 3) silhouette shape-based approaches. The temporal alignment-based approach emphasizes both shape and dynamics and is the most common one. It treats the sequence as a time series and involves three stages of processing. The first stage is the extraction of features such as whole silhouettes [8], [3], principal components of silhouette boundary vector variations [11], silhouette width vectors [6], preshape representation [19], silhouette parts [20], [7], or Fourier descriptors [12]. The gait research group at the University of Southampton (Nixon et al.) has probably experimented with the largest number of possible feature types for recognition. This step also involves some normalization of size to impart some invariance with respect to distance from camera. The second step involves the alignment of sequences of these features, corresponding to the given two sequences to be matched. The alignment process can be based on simple temporal correlation [8], [11], dynamic time warping [19], hidden Markov models [6], [19], [13], phase locked-loops [3], or Fourier analysis [20], [12]. The third aspect is the distance measure used, which can be Euclidean [11], simple dot product-based [8], [6], based on probabilistic models, Procrustes distance [19], or derived based on manifold analysis [21].

The second class of approaches opts for parameters that can be used to characterize gait dynamics, such as stride length, cadence, and stride speed [16]. Sometimes static body parameters, such as the ratio of sizes of various body parts are considered in conjunction with these parameters [16], [17]. However, these approaches have not reported high performances on common databases, partly due to their need for 3D calibration information.

The third class of approaches emphasize the silhouette shape similarity and disregards or underplays temporal

information. One approach involves the transformation of the silhouette sequence into a single image representation. The simplest such transformation is the averaged silhouette [18], [22] or an image representation derived from the width vectors in each frame (Frieze patterns) [23]. Similarity can be based on just the Euclidean distance [22], linear discriminant based distances [18], or symmetric group theoretic distances [23]. The amplitude of the Discrete Fourier transform of the raw silhouette sequence has also been used for recognition [24]. Another way of using shape information preserves individual silhouettes but disregards the sequence ordering and treats the sequences as just a collection of silhouette shapes [15]. Our approach also falls in this category of gait algorithms that emphasize shape over dynamics. However, unlike the approaches that arrive at one representations averaged over all the stances, we use stance specific representations. Like in [15], we do ignore the dynamics between the stances, but, unlike them, we do exploit the temporal ordering of the individual gait stances.

To emphasize the shape aspect of gait, we propose normalizing the gait dynamics based on a population-based generic walking model. Note that we do not normalize the gait on a per-subject basis, but, rather all gaits, are mapped to one model. Fig. 2 depicts the our approach. Given the natural model of gait as involving state transitions, we use a Hidden Markov Model to normalize gait dynamics. In particular, we use a *population* Hidden Markov Model (pHMM), capturing the average dynamics of a set of individuals. The states of this pHMM represent gait stances over one gait cycle and the observations are the silhouettes of the corresponding gait stances. The pHMM can be looked upon as being a *generic* walking model. Note that our use of HMMs has a number of essential differences from the previous applications of HMM in gait recognition [6], [19], [13]. First, we do not have *person specific* HMMs; we use a *population* HMM model, which can be looked upon as a *generic* walking gait model, defined for a set of people. Second, the HMM is *not* used for recognition; it is used just to align the frames of two sequences. Third, temporal dynamics play no role in the similarity computation.

To normalize gait in any given sequence, we use Viterbi decoding of the gait dynamics to map the given sequence stances onto the pHMM states. Then, the stances of each state are averaged to arrive at one dynamics-normalized, averaged, gait cycle of fixed length, equal to the number of states in the pHMM. We compute the biometric similarity between two normalized gait signatures by summing the shape distances between the corresponding gait stances. Shape distances between two silhouettes from the same generic gait stance is computed in the linear discriminant analysis (LDA) space that maximizes the discrimination between persons while minimizing the variations of the same subject under different conditions. To handle noise variations in silhouette shapes that can occur with changing imaging conditions and segmentation thresholds, we structure the distance computation to make it somewhat invariant to dilations and erosions of the silhouettes.

A few words regarding biometrics nomenclature are in order. The term *gallery* is used to refer to the set of templates or sequences stored in the model base. *Probes* are the unknown templates to be identified or verified. In an identification scenario, one is interested in finding a match to a given probe from the whole gallery set, i.e., one-to-many match. In a

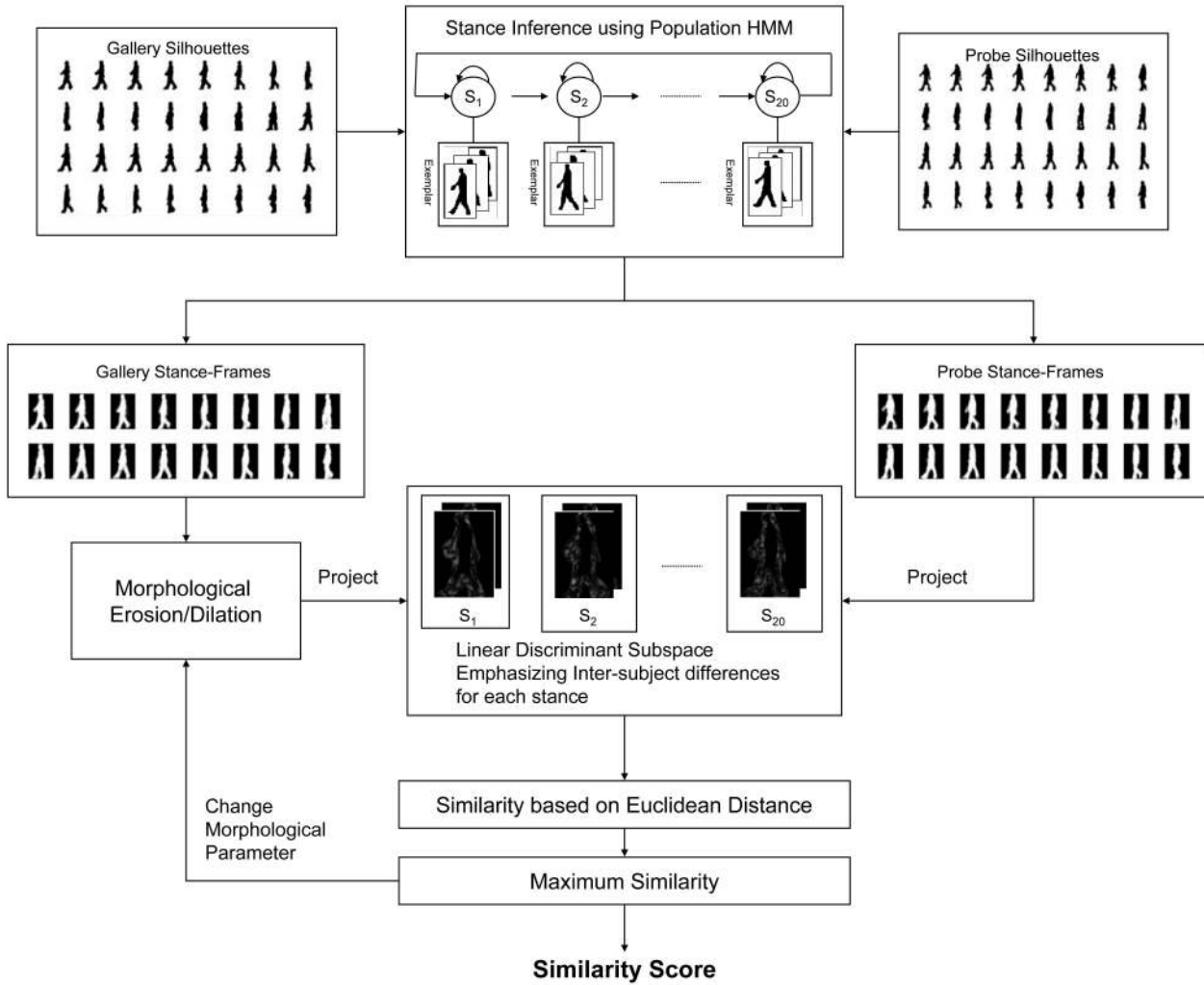


Fig. 2. The gait recognition approach based on gait-dynamics normalization. The different algorithmic modules are shown, along with example inputs and intermediate representations.

verification scenario, one is interested in deciding whether a given probe matches a hypothesized or claimed gallery identity, i.e., one-to-one match. Performance for the identification scenario is captured by the Cumulative Match Characteristic (CMC) [25], which plots identification rates (P_f) within a given rank k . For the verification scenario, the standard Receiver Operator Characteristic (ROC) is used. ROC plots the correct detection rate against the false alarm rate for various choices of the decision threshold.

In the next section, we present the population HMM that we use to normalize gait sequences. This is followed by a discussion of the dynamics normalization process in Section 3. The linear discriminant space-based similarity computation process along with deformation resistant distance measure are described in Sections 4 and 5. In Section 6, we present performance of the proposed approach on three public databases. And, in Section 7, we discuss the implications, conclusions, and possible future research directions.

2 POPULATION HIDDEN MARKOV MODEL (pHMM)

The inputs consist of silhouette sequences, which can be extracted from raw sequences in a number of ways. We compute the silhouettes using the eigenstance reconstruction

model [9], which linearly projects each frame into the eigenstance space corresponding to the mapped state and then reconstructs it. This was shown to significantly reduce the effect of shadows and other segmentation errors. To remove possible bias in recognition due to the use of silhouette height, we normalize the height of the silhouettes to occupy 128 pixels, as in the HumanID gait baseline algorithm [8]. To facilitate the ready comparison of silhouette templates, following the baseline algorithm [8], we also center the silhouettes in each frame along the horizontal direction so that the centerline of the torso is at the middle of the frame. Details of the silhouette extraction, height normalization, and silhouette centering steps can be found in referenced papers. We do not describe it in any more detail here.

Each silhouette sequence, typically consisting of multiple gait cycles, is first aligned to form one dynamics-normalized, averaged gait cycle, over a fixed number of stances. This normalization is accomplished by a *Population Hidden Markov Model (pHMM)* constructed based on a subject population. Like traditional HMM, the Population Hidden Markov Model (pHMM) is specified by the possible states, $q_t \in \{1, \dots, N_s\}$, which represent gait stances, and the triple parameters $\lambda = (A, B, \pi)$, which represent the state transition matrix, observation model, and priors, respectively. The model is



Fig. 3. The top row shows the original images, cropped around the person, for one sequence. The middle row shows the corresponding part-level, manually specified silhouettes. And, the bottom row shows the size-scaled and spatially aligned silhouettes that are used to train the pHMM.

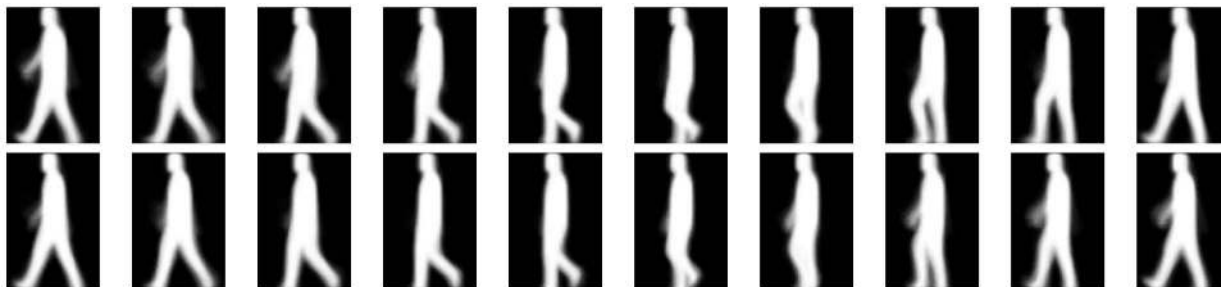


Fig. 4. Average stances in population exemplars for N_s aligned states over a gait cycle.

trained on a set of manually created silhouettes, over one gait cycle, for a set of subjects. Each gait cycle is chosen to begin at the right heel strike phase of the walking cycle through to the next right heel strike. Fig. 3 shows examples of these manual silhouettes. We also size-normalize and spatially align these silhouettes (see the third row of Fig. 3) to reduce the effect of varying distances of the subject from the camera. In our earlier work [9], we first introduced the idea of the population HMM for silhouette reconstruction and cleaning. Here, we use it for a different purpose, i.e., stance alignment. For completeness, here, we summarize the specifics of the model.

2.1 Observation Model

We use the distances of a given observed silhouette from an exemplar set as the observation variables. The exemplar set for each state is computed by K-means clustering of the frames of the training sequences. Of course, for clustering, a distance measure is needed. We define the distance between two vertically scaled and horizontally aligned (see row 3 in Fig. 3) silhouettes, \mathbf{f}_i and \mathbf{f}_j , as function of dot products of the corresponding row-scanned column vectors:

$$S(\mathbf{f}_i, \mathbf{f}_j) = \frac{\mathbf{f}_i^T \mathbf{f}_j}{\mathbf{f}_i^T \mathbf{f}_i + \mathbf{f}_j^T \mathbf{f}_j - \mathbf{f}_i^T \mathbf{f}_j}. \quad (1)$$

Note that, for binary silhouettes, i.e., with pixels values being just 0 or 1, this similarity is the ratio of the pixels in the intersection of the two overlapped silhouettes to the number of pixels in their union and is also commonly known as the Tanimoto similarity measure. One minus this similarity is the Tanimoto distance metric for binary silhouettes; $D(\mathbf{f}_i, \mathbf{f}_j) = 1 - S(\mathbf{f}_i, \mathbf{f}_j)$. For nonbinary silhouettes too, we refer to the above distance (similarity) measure as the Tanimoto distance (similarity).

The exemplar sets are initialized by equally partitioning the frames in one gait cycle into N_s segments. To pick up the asymmetry in gait, i.e., to differentiate stances with left foot forward from those with right foot forward, we use one full cycle (two strides). We group the frames within the j th partition of all people into an exemplar set for the j th gait stance, E_j . Since the gait cycles of the manual silhouettes are aligned, this strategy of corresponding the exemplars from different subjects works.

The exemplars are then refined by K-means clustering with some constraints. Given an initial exemplar set, $E_j^{[0]}$, its frames are iteratively reassigned based on the distance, $D(\mathbf{f}_i, \mathbf{f}_j) = 1 - S(\mathbf{f}_i, \mathbf{f}_j)$:

$$\overline{E}_j^{[k]} = \frac{1}{N} \sum_{\mathbf{f}_i \in E_j^{[k]}} \mathbf{f}_i, \quad (2)$$

$$E_j^{[k+1]} = \left\{ \mathbf{f}_i \mid D(\mathbf{f}_i, \overline{E}_j^{[k]}) < \left(D(\mathbf{f}_i, \overline{E}_{j-1}^{[k]}) \right), D(\mathbf{f}_i, \overline{E}_{j+1}^{[k]}) \right\}. \quad (3)$$

Note the constraint that frames can only be reassigned to neighboring exemplar sets; thus, a frame in E_j can be reassigned to exemplars E_{j-1} or E_{j+1} . In addition, to maintain the transition order among exemplars, a frame can be reassigned to E_j only if it is the next neighbor E_j in terms of stance sequence. To capture the scatters between subjects, we also insist that every exemplar should contain no less than one frame from each sequence. We stop when no more reassignments can be done; about 10 iterations were enough for our experiments. Fig. 4 shows the mean silhouettes, \overline{E}_j , of exemplar sets of the stances over one gait cycle.

Since we have formed N_s exemplars, given an input silhouette frame, \mathbf{f}_t , there are N_s observation variables. The observation model is comprised of a model for each state, $B = \{b_j(\mathbf{f}_t) | j = 1, \dots, N_s\}$, where $b_j(\mathbf{f}_t) = P(\mathbf{f}_t | q_t = j)$, i.e., the conditional probability of the observed silhouette, \mathbf{f}_t , at time t given that the state at time t is j . We choose the observation model to be exponential in terms of the observation variable (the Tanimoto distance, D , between any given silhouette, \mathbf{f}_t , and the mean of the state exemplars, \bar{E}_j).

$$b_j(\mathbf{f}_t) = \frac{1}{\mu_j} e^{-\frac{D(\mathbf{f}_t, \bar{E}_j)}{\mu_j}}. \quad (4)$$

The observation model is thus parameterized by the mean μ_j . The pHMM structure is somewhat similar to that individual HMM structure used in [26] for recognition, but, in our case, it is designed to model gait dynamics over a population. Differences also exist in the observation model and the state definitions; our model takes into account the gait asymmetry between the two strides over a cycle.

2.2 Model Parameter Estimation

We pick equal state priors, i.e., $\pi_i = \frac{1}{N_s}$, since, in practice, any given sequence can begin from any state. However, both the transition matrix and the observation model parameters need to be estimated. Since the exemplar sets have been computed from the given training sequences, we just estimate the observation model parameters for each stance directly from the corresponding exemplars.

$$\mu_j = \frac{\sum_{\mathbf{f}_i \in E_j} D(\mathbf{f}_i, \bar{E}_j)}{|E_j|}. \quad (5)$$

To model the periodic gait motion, we constrain the state transition matrix A , with entries $a(i, j) = P(q_{t+1} = j | q_t = i)$, to represent a cyclical version of the left to right Bakis state transition model over N_s states, allowing for jumps only to the next state. The initial estimate of the transition matrix is also formed from the exemplars and then refined using Levinson's method for training with multiple observation sequences based on iterative Baum-Welch algorithm.

$$a^{[0]}(i, j) = \frac{\sum_k \# \text{ of } \mathbf{f}_{t+1}^k \text{ in } E_j \text{ given } \mathbf{f}_t^k \text{ in } E_i}{|E_i|}. \quad (6)$$

We refer the reader to standard texts such as [27] for details regarding the Levinson's method. Here, we present just the key equations. Let there be K observation sequences, $\{\mathbf{F}^1, \dots, \mathbf{F}^K\}$. However, since, for each training sequence, we have only one gait cycle to retain the cyclical property, we extend each sequence by appending its first frame to the tail: $\mathbf{F}^i = \{\mathbf{f}_0^i, \dots, \mathbf{f}_{N_F}^i, \mathbf{f}_0^i\}$. The length of the extended \mathbf{F}^k is denoted by T_k . The iterative re-estimate of the population transition probabilities, $A^{[n+1]}$, are given by

$$a^{[n+1]}(i, j) = \frac{\sum_{k=1}^K \frac{1}{P_k} \sum_{t=1}^{T_k-1} \alpha_t^k(i) a^{[n]}(i, j) b_j(\mathbf{f}_{t+1}^k) \beta_{t+1}^k(j)}{\sum_{k=1}^K \frac{1}{P_k} \sum_{t=1}^{T_k-1} \alpha_t^k(i) \beta_t^k(i)}, \quad (7)$$

where $P_k = P(\mathbf{F}^k | \lambda)$, the likelihood of k th observation, and the forward and backward probabilities α^k and β^k , are arrived at by induction as follows:

$$\begin{aligned} \alpha_t^k(j) &= P(\mathbf{f}_1^k, \dots, \mathbf{f}_t^k, q_t = j | \lambda) \\ &= b_j(\mathbf{f}_1^k) / N_s \\ &\quad \text{for } t = 1 \text{ and } 1 \leq j \leq N_s \\ &= \sum_{i=1}^{N_s} \alpha_{t-1}^k(i) a^{[n]}(i, j) b_j(\mathbf{f}_t^k) \\ &\quad \text{for } 2 \leq t \leq T_k - 1 \text{ and } 1 \leq j \leq N_s, \end{aligned} \quad (8)$$

$$P_k = P(\mathbf{F}^k | \lambda) = \sum_{j=1}^{N_s} \alpha_{T_k}(j), \quad (9)$$

$$\begin{aligned} \beta_t^k(i) &= P(\mathbf{f}_{t+1}^k, \dots, \mathbf{f}_{T_k}^k | q_t = i, \lambda) \\ &= 1 \quad \text{for } t = T_k \text{ and } 1 \leq i \leq N_s \\ &= \sum_{j=1}^{N_s} a^{[n]}(i, j) b_j(\mathbf{f}_{t+1}^k) \beta_{t+1}^k(j) \\ &\quad \text{for } t = T_k - 1, \dots, 1 \text{ and } 1 \leq i \leq N_s. \end{aligned} \quad (10)$$

Equations (7)-(10) represent the generalization of the Baum-Welch equations for multiple observations and need to be iterated over until the likelihood of the given observations are maximized. The learned transition matrix emphasizes the transitions to forward states, manifesting as high values along the first upper diagonal. We also found high values at the antidiagonal corner, which is because we adopt a cyclical Bakis model.

2.3 Model Size Determination

We determine the number of states, N_s , based on the Akaike Information Criterion (AIC) [28], which take both the goodness of fit and generalizability into account:

$$\text{AIC} = -2 \sum_{k=1}^K \log_2 P(\mathbf{F}^k | \lambda) + 2N_{para}, \quad (11)$$

where (λ) is the estimated pHMM, K is the number of training sequences, and N_{para} is the number of estimated parameters of the model. The estimated parameters include the N_s^2 transition probabilities and the N_s parameters in the observation model. The variation of AIC with the number of states is shown in Fig. 5 for two different training sets of 71 subjects, one over grass walking surface and the other over concrete walking surface. Notice the similar nature of the trend for the two conditions. Based on these plot, we choose the round figure of 20 states as being fairly optimal for both the sets of sequences. It is better to err toward a larger number of states so as to retain the shape variations among different individuals. Given that the "key-stances" in gait are few, this number of states might appear to be high. However, one should note that HMM states are not meant to be just "key-stances," but rather to provide to a statistically good sampling of the trajectory in the "stance" space. Key-stances can be looked upon as some sort of extremal points on this trajectory. Also, since we are building one HMM to capture variation across a population, we would need more states than per-person HMMs.

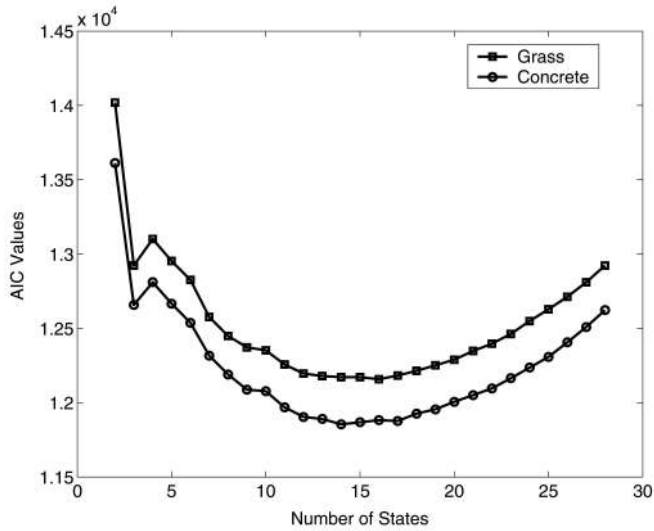


Fig. 5. Variation of AIC with number of states, for pHMMs constructed using two different training sets of 71 subjects; one for a grass walking surface and the other for a concrete walking surface.

3 DYNAMICS NORMALIZED GAIT CYCLE

Using the pHMM, we normalize the dynamics for any given gait sequence, $F = \{f_1, \dots, f_N\}$, by first estimating the stance state for each frame and then averaging the frames mapped to each state to produce one, dynamics-normalized, gait cycle over N_s frames, denoted by $F_{DN} = \{g_1, \dots, g_{N_s}\}$. The dynamics normalized gait cycle is computed by averaging frames mapped to the same state. We refer to this averaged representation for each stance, g_i , as the *stance-frame*. The stance estimation of each frame is based on the dynamic programming-based Viterbi algorithm [27], which returns the most likely state assignment for each frame. To reduce the combinatorics of this assignment process, we partition an input sequence into subsequences of roughly one gait cycle length, which we can easily estimate from the periodic variation in the number of foreground pixels in the bottom half of the silhouettes. This strategy for estimating the gait period was shown to be rather effective in the HumanID Gait Challenge problem [8]. Note that the subsequences can start from any stance because of the cyclical nature of the HMM model.

Fig. 6 shows some stance-frames for one subject under different conditions. Notice that the stance-frames for the same stance are similar across different sequences, which indicates that silhouette-to-stance matching is correctly estimated by the Viterbi algorithm. Also, note the differences in the “width” of the silhouettes for the same stance, but across different conditions. This has to do with the silhouette detection algorithm that used the same set of parameters across conditions, some with different backgrounds, resulting in oversegmentation in some cases. We will take care of this variation, which is inevitable for any silhouette detection algorithm, during similarity computation.

4 STANCE SHAPE DISCRIMINANT SPACE

A dynamics normalized gait cycle consists of a fixed number of stance frames, which simplifies the similarity computation

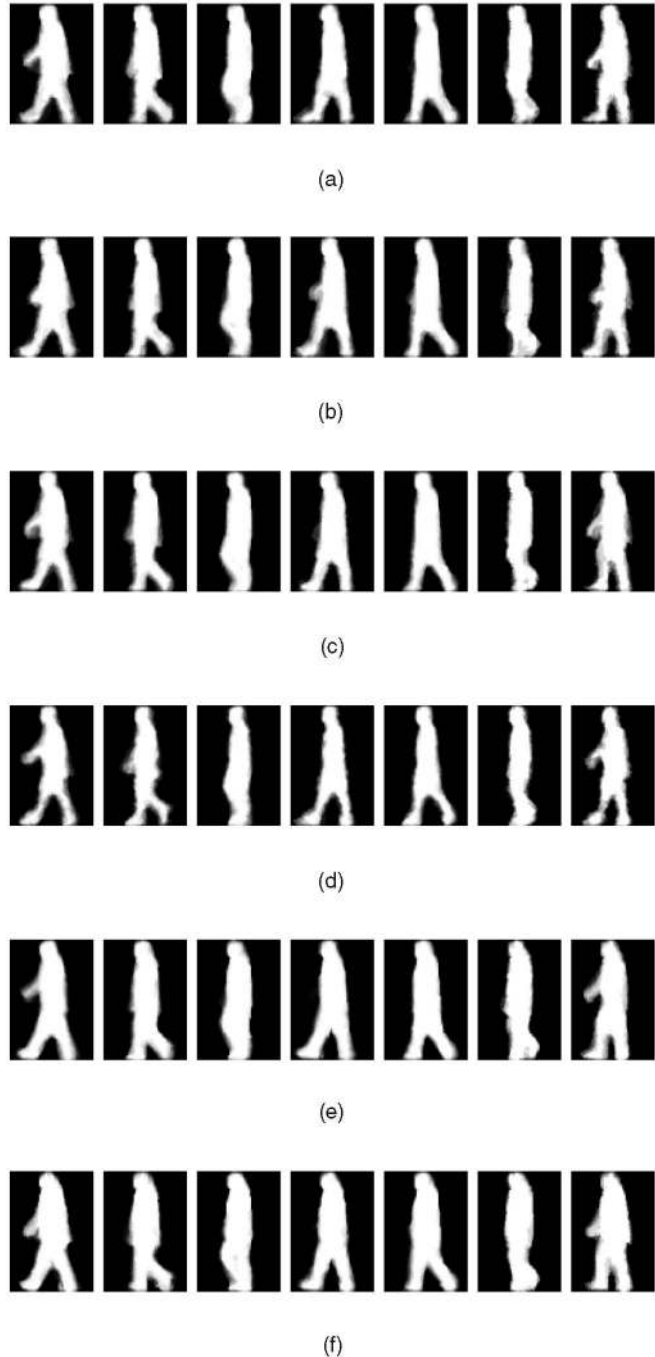


Fig. 6. Example of dynamics-normalized stance-frames from one subject in the (a) gallery and the corresponding stances in the probes corresponding to changes in (b) view, (c) shoe-type, (d) surface, (e) carrying condition, and (f) time (six months).

between two given sequences. A separate alignment process is not needed. We can simply consider the distances between the corresponding stance-frames. Instead of simple Euclidean distances between stance-frames, we compute distances in the Linear Discriminant Analysis (LDA) Space, designed to maximize the differences between frames from different subjects and to minimize the distances between frames from the same subject under different conditions. In the final section of this paper, we present results to demonstrate that LDA does increase recognition performance over simple Euclidean distances between stance frames.

However, instead of pure LDA, we used the PCA + LDA formulation that was advocated by Belhumeur et al. [29] so as to address the singularity issues. Singularity arises because of the large feature vector size compared to the size of the training set. Consider building pure LDA spaces for *each* of the N_s stances. The set of individuals form the classes, (I_1, I_2, \dots, I_c) . For each individual, I_k , the stance-frames, \mathbf{g}_s^i (s is the stance index), under various different conditions form the samples for that class. The between-class scatter matrix for the s stance is

$$S_B = \sum_{i=1}^c N_i (\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T, \quad (12)$$

the within-class scatter matrix is

$$S_W = \sum_{i=1}^c \sum_{\mathbf{g}_s^i \in I_i} (\mathbf{g}_s^i - \boldsymbol{\mu}_i)(\mathbf{g}_s^i - \boldsymbol{\mu}_i)^T, \quad (13)$$

and the total scatter matrix is

$$S_T = \sum_{i=1}^N (\mathbf{g}_s^i - \boldsymbol{\mu})(\mathbf{g}_s^i - \boldsymbol{\mu})^T, \quad (14)$$

where $\boldsymbol{\mu}_i$ is the mean vector of class i , $\boldsymbol{\mu}$ is the mean vector of all samples, and N is the total sample number. If S_W is nonsingular, then the optimal discriminating space \mathbf{V}_{opt} for classification can be simply computed as

$$\mathbf{V}_{opt} = \arg \max_{\mathbf{V}} \frac{|\mathbf{V}^T S_B \mathbf{V}|}{|\mathbf{V}^T S_W \mathbf{V}|}. \quad (15)$$

Specifically, $\mathbf{V}_{opt} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\}$, the set of generalized eigenvectors of S_B and S_W corresponding to the m largest eigenvalues $(\lambda_1, \lambda_2, \dots, \lambda_m)$, i.e.,

$$S_B \mathbf{v}_i = \lambda_i S_W \mathbf{v}_i, \quad i = 1, 2, \dots, m.$$

However, the rank of S_w is no more than $N - c$, where N is the total number of subjects in training set and c is the number of classes. If the dimension of S_w , which is determined by the size of the feature vector, is less than number of samples, $N - c$, then S_w will not be singular. But, our feature vector is the row-scanned image and, so, its size is the product of the height and width of the image. This size, which is on the order of 10,000, is more than $N - c$, which is on the order of 100. So, S_w is singular for most training set sizes. Of course, for training set sizes larger than the number of pixels in the image S_w will not be singular.

One solution is to project the within-class and between-class scatter matrices into a lower dimension space, i.e., reducing the feature vector size and rendering S_W is nonsingular. The PCA (Principal Component Analysis) can be employed to reduce the dimension [29].

$$\mathbf{V}'_{opt} = \mathbf{V}_{PCA} \mathbf{V}_{LDA}, \quad (16)$$

where

$$\begin{aligned} \mathbf{V}_{PCA} &= \arg \max_{\mathbf{V}} |\mathbf{V}^T S_T \mathbf{V}|, \\ \mathbf{V}_{LDA} &= \arg \max_{\mathbf{V}} \frac{|\mathbf{V}^T \mathbf{V}_{PCA}^T S_B \mathbf{V}_{PCA} \mathbf{V}|}{|\mathbf{V}^T \mathbf{V}_{PCA}^T S_W \mathbf{V}_{PCA} \mathbf{V}|}, \end{aligned}$$

and \mathbf{V}_{PCA} should keep no more than the largest $N - c$ principal components so that the corresponding S_W is nonsingular.

4.1 Similarity Expression

For each stance, s , in the dynamics-normalized gait representation, we create \mathbf{V}^s_{PCA} to model only 90 percent energy in the corresponding total scatter matrix S_T . We have found the number of eigenvectors thus needed is much less than $N - c$. The subsequent \mathbf{V}^s_{LDA} space consist of $c - 1$ nonzero generalized eigenvectors of within subjects and between subjects scatter matrix for the k th stance. Given two dynamics normalized sequences, \mathbf{F}^a_{DN} and \mathbf{F}^b_{DN} , we compute the distance by first projecting each stance-frame, \mathbf{g}_k^a into the corresponding \mathbf{V}^s_{LDA} space. This negated sum of the Euclidean distances in these LDA stance spaces is a similarity measure, $S(\mathbf{F}^a_{DN}, \mathbf{F}^b_{DN})$. More specifically,

$$\begin{aligned} S(\mathbf{F}^a_{DN}, \mathbf{F}^b_{DN}) &= - \sum_{k=1}^{N_s} \|\mathbf{g}_k^{aT} \mathbf{V}^s_{LDA} - \mathbf{g}_k^{bT} \mathbf{V}^s_{LDA}\| \\ &= - \sum_{k=1}^{N_s} (\mathbf{g}_k^a - \mathbf{g}_k^b)^T \mathbf{V}^s_{LDA} (\mathbf{V}^s_{LDA})^T (\mathbf{g}_k^a - \mathbf{g}_k^b). \end{aligned} \quad (17)$$

5 SIMILARITY UNDER SILHOUETTE DEFORMATIONS

Similarity computation measure needs to be augmented to handle imaging and low-level processing effects such as that displayed in Fig. 6. Notice that the “thickness” of the stance-frame for the same person varies across different conditions. For instance, there were changes associated with change in surface and time. This type of variation arises because of variabilities of low-level silhouette detection processes, induced by changes in the background statistics. These are hard to completely eliminate. Accepting this constraint, we modify our similarity computation to be somewhat robust with respect to changes in overall silhouette “thicknesses.” The new similarity, S_{WN} , is the maximum possible similarity over possible morphological deformations of the stack-frames in one of the sequences. The morphological deformations of erosion and dilation model the possible variations in thicknesses. Specifically,

$$\begin{aligned} S_{WN}(\mathbf{F}^a_{DN}, \mathbf{F}^b_{DN}) &= \sum_{k=1}^{N_s} \arg \max_{m \in \{-M, \dots, M\}} (Mor(m, \mathbf{g}_k^a) - \mathbf{g}_k^b)^T \\ &\quad \mathbf{V}^s_{LDA} (\mathbf{V}^s_{LDA})^T (Mor(m, \mathbf{g}_k^a) - \mathbf{g}_k^b), \end{aligned} \quad (18)$$

where

$$Mor(m, \mathbf{g}_s^i) = \begin{cases} Dilate(m, \mathbf{g}_s^i) & \text{if } m \geq 0 \\ Erode(m, \mathbf{g}_s^i) & \text{if } m < 0. \end{cases} \quad (19)$$

The *Erode* and *Dilate* are gray-level morphological operations, as implemented in Matlab.

Fig. 7 shows the morphologically processed stance-frames in the *gallery*, corresponding to the probes in Fig. 6, that maximized the overall similarity. We see that eroded forms of some gallery stance-frames, e.g., in the first and fourth columns, are more similar to the surface-probe shown in Fig. 6 than the original gallery stance-frames. Another example is the stance-frames shown in the third, sixth, and seventh

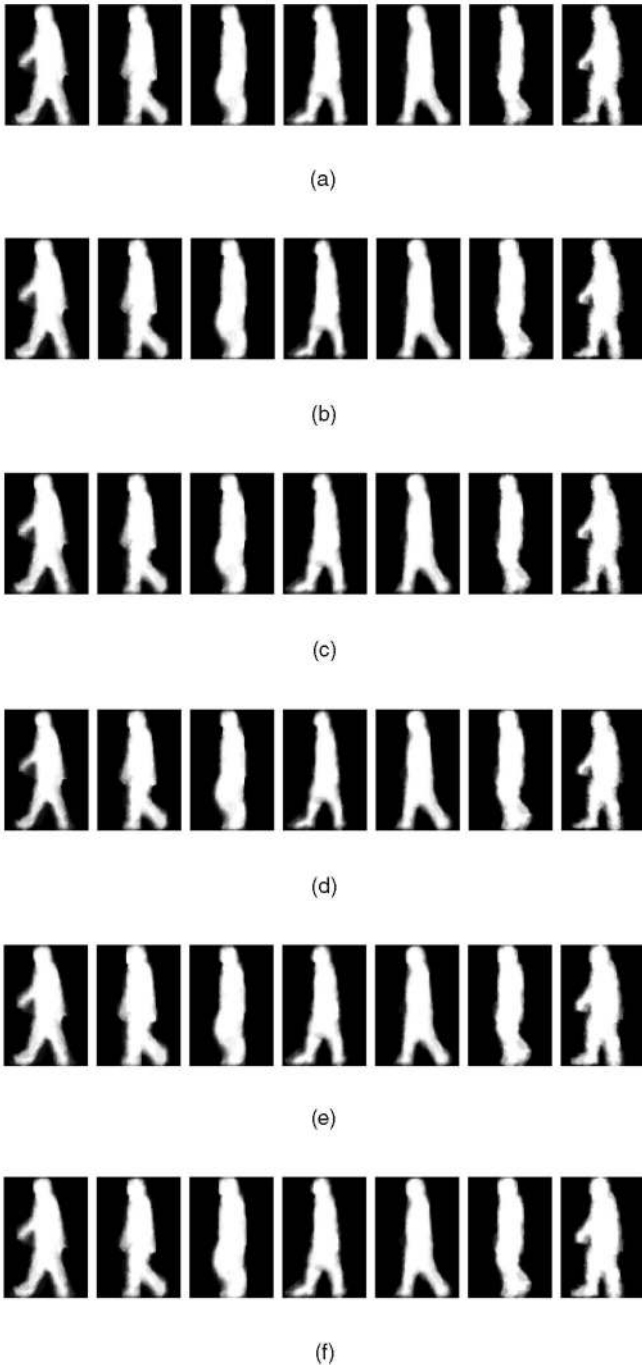


Fig. 7. Examples of the average stances in the gallery set (a) before and after morphological operation for best shape match to (b) Probe-View, (c) Probe-Shoe, (d) Probe-Surface, (e) Probe-Briefcase, and (f) Probe-Time in Fig. 6.

columns for the time-probe. This is consistent with our initial observation that the silhouettes in the surface-probe appear to be thinner than the gallery while those in time-probe appear to be thicker. The new deformation invariant similarity measure helps us handle such cases.

6 EXPERIMENTS AND ANALYSIS

We present results on three different publicly available data sets: the HumanID gait challenge data set, the UMD outdoor data set, and the CMU indoor MoBo data set. We

study the ability of the proposed algorithm to improve performance for matching across surfaces, time, carrying condition, and walking speed variations. We also present results for varying degrees of separation between training and test sets demonstrating generalizability across individuals, data collection sites, and camera configurations. We start by briefly describing the data sets.

6.1 USF/NIST HumanID Gait Challenge Database

The USF/NIST HumanID gait challenge database was formulated to facilitate objective, quantitative measurement of gait research progress on a large data set. It is presently the largest available data set. Fig. 8 shows some sample frames. It was collected outdoors and each person in the data set was studied under combination of as many as five conditions. The conditions are:

1. two camera angles,
2. two shoe types,
3. two surfaces (grass and concrete),
4. with and without carrying a brief case, and
5. two different dates six months apart, May and November.

An attempt was made to acquire a person's gait in all possible combinations and there are up to 32 sequences for some people, so that a full data set can be partitioned into 32 subsets, one for each combination of the five covariates. Comparisons between these subsets are used to set up challenge experiments; more on this later. The full data set consists of 1,870 sequences from 122 individuals. This data set is significantly larger than other data sets and is also unique in the number of covariates exercised. It is the only data set to include walking on a grass surface.

6.2 UMD Gait Database

There are two UMD gait data sets: data set-1 consists of walking sequences of 25 subjects and data set-2 contains walking sequences of 55 subjects walking along a T-shape pathway. In this paper, we use the larger one: data set-2, taken outdoor by two surveillance cameras (Philips G3 Enviro-Dome camera system) at a height of 4.57 meters. Fig. 9 shows one sample frame. Each video sequence has approximately 10 gait cycles, viewed frontally and sideways. The database is diverse in terms of gender, age, and ethnicity. Moreover, like the gait challenge database, data collected on different days differ with respect to clothing as well. There are two significant differences in imaging protocol with gait challenge data set: 1) The camera sample rate of the UMD data is 20 frame per second (f/s) but that of the gait challenge data is 30 f/s and 2) the camera was setup at 4.57 meters from the ground for the UMD data, but it was 1.65 meters high for the gait challenge data.

6.3 CMU Mobo Database

The CMU Mobo data set [5] consists of sequences from 25 subjects walking on a treadmill, positioned in the middle of a room. Fig. 10 shows some sample frames. Each subject is recorded performing four different types of walking: slow walk (2.06 miles/hr), fast walk (2.82 miles/hr), slow walk holding a ball, and walk on an inclined plane. Each sequence is 11 seconds long, recorded at 30 frames per second. Six cameras were set up to take images from side view, diagonal view, frontal view, and back view.



Fig. 8. Samples from the HumanID gait challenge data set: subject walking on grass (a) along the frontal half of the elliptical path and (b) along the back half of the elliptical path.

6.4 Training and Test Sets

In biometrics, the test set is comprised of two subsets. One subset of sequences are defined as the gallery, representing prestored templates, and another subset of sequences are defined to be the probes, representing templates to be verified or identified. It is usually acceptable to use the gallery sequences for training; however, ideally we would like to use a separate set of training sequences. The separation can be in terms of a third set of sequences for the same set of subjects or, more strongly, in terms of sequences from a separate set of subjects. However, for recognition algorithms that employ per-subject models, this strong separation in terms of subjects is not possible. Fortunately, since our algorithm uses a generic model, it can be built using data from a completely separate set of subjects. The LDA space can also be built using a separate set of subjects. This allows us to have complete separation of test and train sets, even in terms of subjects. However, in our experiments, we do show results with training sets with varying degrees of separation.

All training was done using *subsets* chosen from the 32 subsets of the HumanID Gait Challenge data. Recall that the subsets corresponds to the possible combinations of the

five covariates, with two possible choices each. The pHMM was trained using the *manual* silhouettes over *one* gait cycle from 71 subjects in the HumanID gait challenge data. Specifically, we choose 71 manual silhouette sequences corresponding to subjects walking on grass, viewed from the right camera, in the May collection. As we will see later, the real data corresponding to these manual silhouette form the *gallery* of the experiments defined in the gait challenge data set. None of the data from the probes were used for training the pHMM. In terms of experimental protocols, this offers us acceptable separation of train and test conditions for the experiments on the gait challenge data set. For experiments with the UMD and CMU data sets since we do not retrain the pHMM, there is *complete* separation of train and test in terms of subjects.

To create the linear discriminant stance spaces, we also need a training set, comprised of stance-frame samples from different subjects under different conditions. For this, we used *subsets* of the HumanID gait challenge data to construct two different training data sets to allow us to experiment with different level of differences between train and test sets. The first training set consists of data from 33 subjects, with six



Fig. 9. Sample frame from the UMD gait database in which subjects walked along a T-shape path in front of two orthogonally placed cameras. We used the fronto-parallel view portion for recognition.

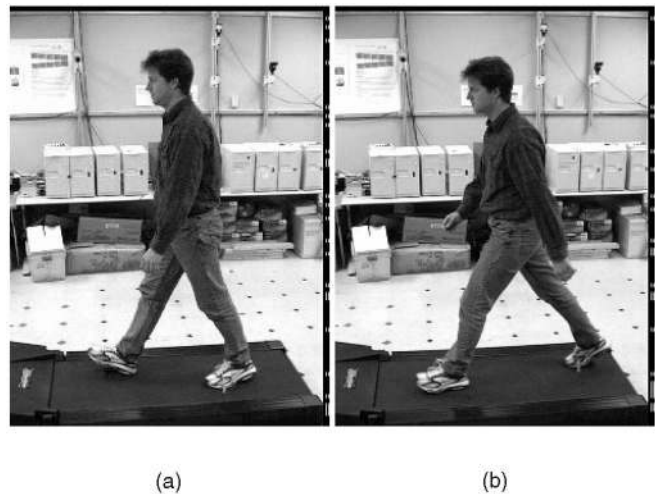


Fig. 10. Samples of the CMU gait database walking on a treadmill in the middle of a room under the condition of (a) slow walk and (b) fast walk.

TABLE 1
The Gallery and Probe Set Specifications for Each of the Gait Challenge Experiments

Exp.	Probe (Surface, Shoe, View, Carry, Time) (C/G, A/B, L/R, NB/BF, time)	Number of Subjects	Difference
A	(G, A, L, NB, M + N ₁)	122	V ¹
B	(G, B, R, NB, M + N ₁)	54	S ²
C	(G, B, L, NB, M + N ₁)	54	S+V
D	(C, A, R, NB, M + N ₁)	121	F ³
E	(C, B, R, NB, M + N ₁)	60	F+S
F	(C, A, L, NB, M + N ₁)	121	F+V
G	(C, B, L, NB, M + N ₁)	60	F+S+V
H	(G, A, R, BF, M + N ₁)	120	B ⁴
I	(G, B, R, BF, M + N ₁)	60	S+B
J	(G, A, L, BF, M + N ₁)	120	V+B
K	(G, A/B, R, NB, N ₂)	33	T ⁵ +S+C ⁶
L	(C, A/B, R, NB, N ₂)	33	F+T+S+C

¹View, ²shoe, ³surface, ⁴carry, ⁵time, and ⁶clothing. The covariates are coded as follows: C—concrete surface, G—grass surface, A—first shoe type, B—second shoe type, BF—carrying a briefcase, NB—no briefcase, M—data collected in May, N₁—new subjects in November data, and N₂—repeat subjects in November. The gallery for all of the experiments is (G, A, R, NB, M + N₁) and consists of 122 individuals.

sequences per subject, corresponding to the various combinations of changes in four covariates: view-point, surface-type, carry-condition, and time. Specifically, we used subsets (G, A, R, NB, M), (C, A, R, BF, M), (C, A, L, BF, M), (C, A, L, NB, N₁), (G, A, L, NB, N₁), and (G, A, R, BF, N₁), where the covariates are coded as follows: C—concrete surface, G—grass surface, A—first shoe type, B—second shoe type, BF—carrying a briefcase, NB—no briefcase, M—data collected in May, N₁—new subjects in November data, and N₂—repeat subjects in November. Note that we chose only those subsets for training that were not part of any probe sets in the HumanID Gait Challenge experiments (see Table 1).

The second training set for the LDA consists of automated silhouette data from 51 subjects collected in November, with no overlap with the May subjects. This training set allows us to complete separation of train and test on the HumanID Gait Challenge data set itself. The corresponding test set consists of sequences from the May collection. In the train set, for each subject, we had eight sequences corresponding to the various combinations of two-values of view point, surface, and carrying conditions. We did not include shoe-variation in the training sets as its inclusion reduced the number of common subjects for each combination of conditions. This is not of much concern because, as reported results on the Gait Challenge problem indicate, the impact of shoe on gait recognition is the lowest [8].

We used test sequences with varying degrees of differences with the above training sets in terms of the subjects. The first test set consists of sequences from 122 subjects, including the training subjects in the first training set for the LDA and those use the pHMM. The second test set consists of sequences from the Gait Challenge data set collected in May and does not include *any* subject from the second training set. The third and fourth test data sets correspond to the UMD and CMU data sets, respectively. These sets not only have separation in terms of subjects, but were also collected at different sites, with different viewing geometry and cameras than the training set.

TABLE 2

Performance of the Dynamics-Normalized Gait Recognition Algorithm for the HumanID Gait Challenge Experiments (with 122 Subjects) in Terms of the Identification Rate P_I at Ranks 1 and 5 and the Verification Rate P_V for False Alarm Rates of 1 Percent and 10 Percent

Exp.	P_I at		P_V at	
	Rank 1	Rank 5	$P_{F=1\%}$	$P_{F=10\%}$
A	85%	96%	93%	98%
B	89%	94%	94%	98%
C	72%	89%	80%	94%
D	57%	85%	68%	96%
E	66%	81%	62%	90%
F	46%	68%	53%	86%
G	41%	69%	43%	79%
H	83%	96%	91%	99%
I	79%	95%	86%	97%
J	52%	79%	58%	92%
K	15%	46%	27%	61%
L	24%	39%	24%	46%

6.5 Gait Challenge Problems: View, Shoe, Surface, Carry, and Time

Along with the data set, the gait challenge problem comes with the definition of a set of 12 challenge experiments (A through L), spanning different levels of difficulty. This provides a common benchmark to compare performance with other algorithms. In this section, we present results for these experiments. The experiments are designed to investigate the effect on performance of the five factors, i.e., change in viewing angle, change in shoe type, change in walking surfaces (concrete and grass), carrying or not carrying a briefcase, and temporal differences. The results from the 12 experiments provide an ordering of the difficulty of the experiments. The signatures are the video sequences of gait. To allow for a comparison among a set of experiments and limit the total number of experiments, the gallery set was fixed as the control. Then, 12 probe sets were created to examine the effects of different covariates on performance. The gallery consists of sequences with the following covariates: Grass, Shoe Type A, Right Camera, No Briefcase, and collected in May along with those from the *new* subjects from November. This set was selected as the gallery because it was one of the largest for a given set of covariates. The structure of the 12 probe sets is listed in Table 1. The last two experiments study the impact of time. The time covariate implicitly includes a change of shoe and clothing because we did not require subjects to wear the same clothes or shoes in both data collections. Because of the implicit change of shoe, we assume that a different set of shoes were used in the May and November data collections. This is noted in Table 1 by A/B for shoe type in experiments K and L.

Table 2 lists the identification performance in terms of P_I at rank 1 and rank 5, and verification rate at false alarm rate of 1 percent and 10 percent, respectively. Fig. 11 compares the top rank identification performance achieved with those reported in the literature for the full Gait Challenge problem. Specifically, we compare with the baseline algorithm that came with gait challenge problem [8], UMD's HMM-based recognition strategy [6], and UCR's gait energy + learning-based strategy [18]. We see that the new dynamics-normalized algorithm achieves the best performance in most experiments. It is slightly low for experiments A (view) and C (view + shoe). Of particular interest is the dramatic

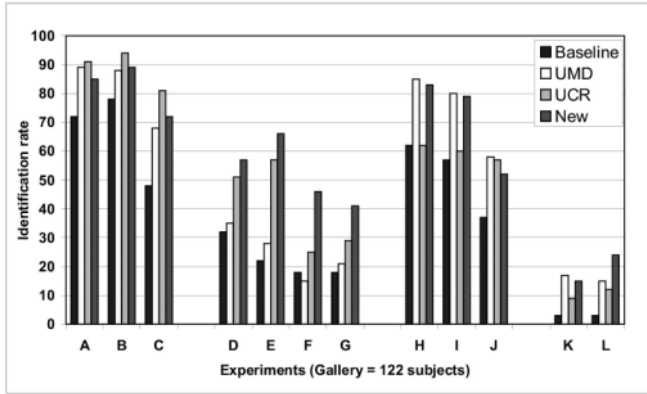


Fig. 11. Top rank recognition rate comparison between the dynamics-normalized (new) gait recognition algorithm with results reported by other algorithms: the baseline algorithm [8], UMD's HMM-based algorithm [6], and UCR's algorithm [18], for the full HumanID data set (122 subjects).

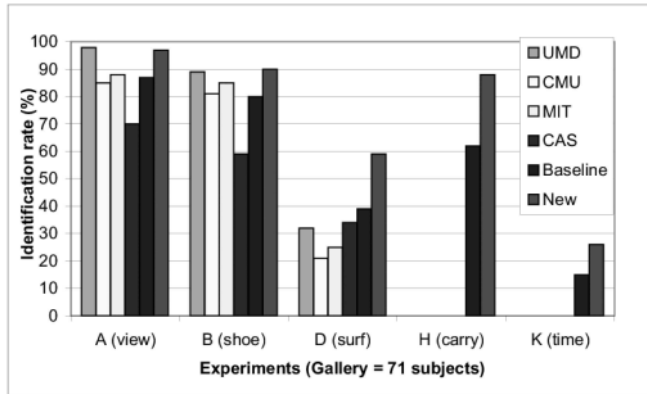


Fig. 12. Summary of the top rank recognition for experiments A (viewpoint), B (shoe-type), D (surface), H (Carry), and K (Time) for first release of HumanID Gait Challenge data set (71 subjects in the May collection). The algorithms are based on: HMM-based algorithm (UMD) [19], silhouette shape clustering (CMU) [15], HMM (MIT) [13], body shape (CAS) [11], and baseline (USF) [8].

improvements for experiments involving surface change (experiments D, E, F, G) and carrying condition change (experiments H, I, J). There is also some increase in performance for the hardest experiments involving six months time-difference (experiments K and L). Note the UMD data set-2 is perhaps the better data set for considering the time-covariate, consisting of data from 55 subjects and time variations over a week. We report performance on this data set in a later section. It is also worthwhile noting the differences in the train sets among the algorithms. The baseline algorithm does not need training. UCR and UMD's algorithm were trained on the gallery. And, the dynamic-normalized algorithm, as spelled out earlier, uses other subsets (nonprobe), in addition to the gallery, to train.

We also experimented with the gait challenge data set with complete separation of training and test sets in terms of the subjects; no subject used for training was part of any probe. We used the second train set, discussed earlier, consisting of sequences from subjects in the November collection who were *not* in the May collection. The test set consisted of sequences from 71 subjects in the May collection and sequences from the repeat subjects in November. Of course, the experiment specifications in Table 1 have to be reduced to include just the May sequences (M) and exclude all the November sequences with N_1 tag. Incidentally, this corresponds to the experiments

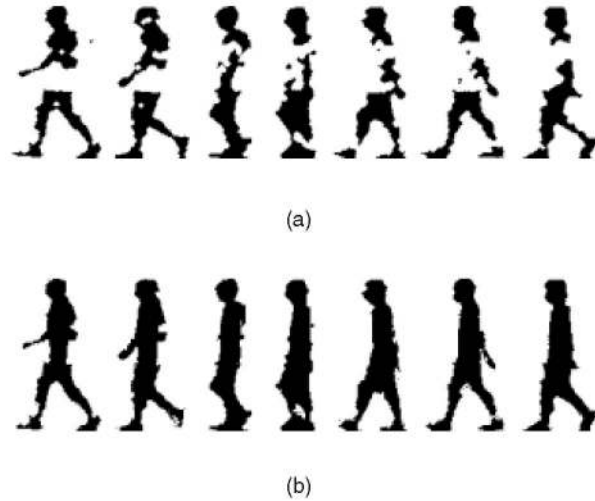


Fig. 13. (a) Sample raw silhouettes from the UMD data set with (b) corresponding cleaned silhouettes based on the Eigenstance model [9].

for the first release of the gait challenge problem on which more groups have reported performance than for the full data set [8]. Fig. 12 shows the performances for the five key experiments: A-view, B-shoe, D-surface, H-carrying, and K-time, based on the new algorithm, as well as those reported by others. We see that dynamics-normalization significantly improves performance, even with complete separation of the training and test sets in terms of subjects.

6.6 UMD Database: Time

The UMD data set-2 offers us an opportunity to test gait recognition with short term (days) time differences for 55 subjects. Specifically, we use the UMD specifications of *experiment 1 for data set-2*, which compares sequences taken on different days. For a more detailed description of the data set and the experiment specification, please refer to <http://degas.umiacs.umd.edu/Hid/data.html>. We chose only the fronto-parallel sections for this experiment. As for all our experiments, the raw silhouettes that were provided along with the data were cleaned using the eigenstance reconstruction model, defended elsewhere in [9], which linearly projects each frame into corresponding eigen-stance space and then reconstructs it. Fig. 13 shows an example of the improvement in the silhouette quality due to this process.

There was no separate training of the gait recognition algorithm on this data set. We use the version trained on the gait challenge data. Fig. 14 shows the top rank identification rate for the new dynamics-normalized gait recognition algorithm at 81 percent is a big improvement over the 71 percent rate of the baseline algorithm, 55 percent of UMD's HMM-based algorithm [6], and 70 percent of MIT's algorithm [30]. This improvement can be attributed to the dynamic-normalization process that removed dynamics variabilities over time. Note that we could not find the rank 1 identification rate of CMU's key stance approach [14]; the lowest rank reported in the paper corresponds to the top 1 percent, which computes to roughly rank 2.

6.7 CMU MoBo Database: Speed

The CMU MoBo database, collected indoors on a treadmill, supports the study of gait recognition variation with respect to walking speed. Specifically, we use experiment 3.1,

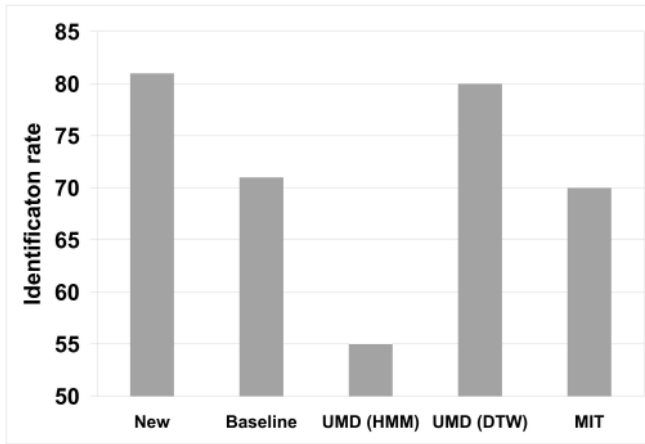


Fig. 14. The top rank identification rates on the UMD data set (experiment 1, 55 subjects) for the dynamics-normalized gait recognition algorithm, baseline algorithm, UMDs HMM-based algorithm [6] (44 subjects), UMD's DTW-based approach [31] (44 subjects), and MIT's algorithm [30].

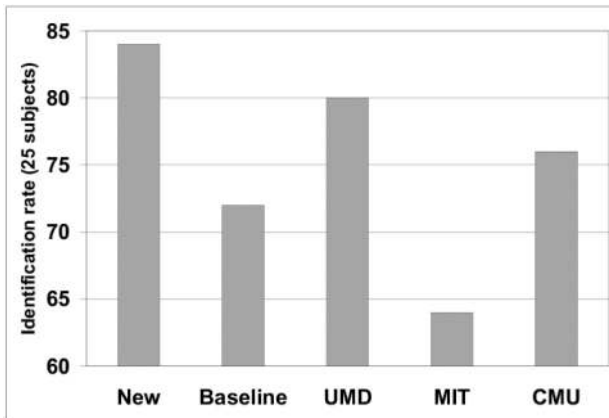
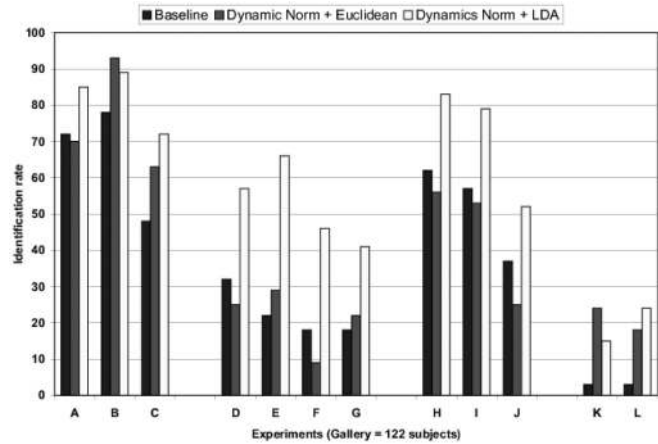


Fig. 15. The top rank identification rate on the CMU Mobo data set (experiment 3.1, 25 subjects) for the dynamics normalized algorithm, baseline algorithm, UMD algorithm [19], MIT algorithm [30], and CMU algorithm [14].

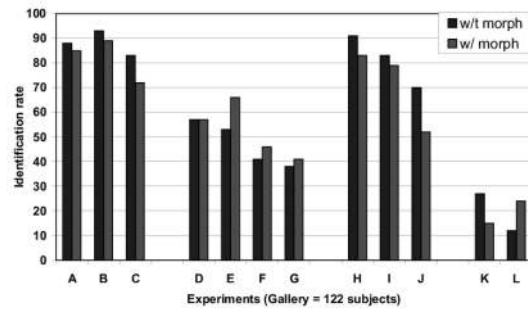
defined by CMU, to test gait recognition across different speeds, as viewed from the side. This is ideal for testing our dynamics-normalization scheme and benchmarking it with respect to performances of other gait recognition approaches that do not normalize dynamics. Like the experiments with the UMD data set, we *did not* retrain the dynamics-normalization model on this data set. As Fig. 15 shows, the performance with dynamics-normalization is high, when compared with reported performances of four other algorithms: baseline [8], UMD [19], MIT [30], and CMU [14]. Given the small size of the Mobo data set, we have included performances of only those algorithms that have also published performances on another larger data set of at least 50 subjects.

7 CONCLUSIONS AND DISCUSSIONS

We presented a dynamics-normalized gait recognition algorithm built around a population Hidden Markov Model (pHMM), coupled with a Linear Discriminant Analysis based shape space, emphasizing differences in stance shapes between subjects and suppressing differences for the same subject under different conditions. The similarity



(a)



(b)

Fig. 16. Performance with variation of the proposed dynamics normalized algorithm. The top rank identification rates on the 12 gait challenge experiments on the HumanID Gait data set for (a) the baseline algorithm, dynamics normalization + Euclidean, and the dynamics normalization + LDA, and (b) with and without silhouette deformation during distance computation.

computation in this space was designed to be robust with respect to “thickening” or “thinning” of silhouettes due to variations in low-level thresholds. Unlike other HMM-based gait algorithms [13], [19] that uses HMMs for *recognition*, we do not use it for recognition, but rather for *dynamics-normalization*. Consequently, in contrast to other HMM-based gait recognition algorithms that build one HMM for each person, we use *one* population HMM.

Based on extensive experimentation on multiple, publicly available databases (HumanID Gait Challenge, UMD, and CMU-Mobo), we can assert that dynamics-normalization improves overall gait recognition performance, especially when comparing across surface, carrying condition, time, and different speeds. The approach is also not dependent on the training set choice. It generalizes well not only across different subjects, but also across different data sets with varying imaging contexts. We attribute this significant improvement to gait dynamics-normalization. The other two components: The LDA stance shape space and the morphological operation-based distance computation also improve performance, but the former has more impact than the latter. Fig. 16a shows the identification rates for the 12 experiments on the HumanID gait data set with and without LDA, as compared to the baseline algorithm, which emphasizes both shape and dynamics. Notice that, although

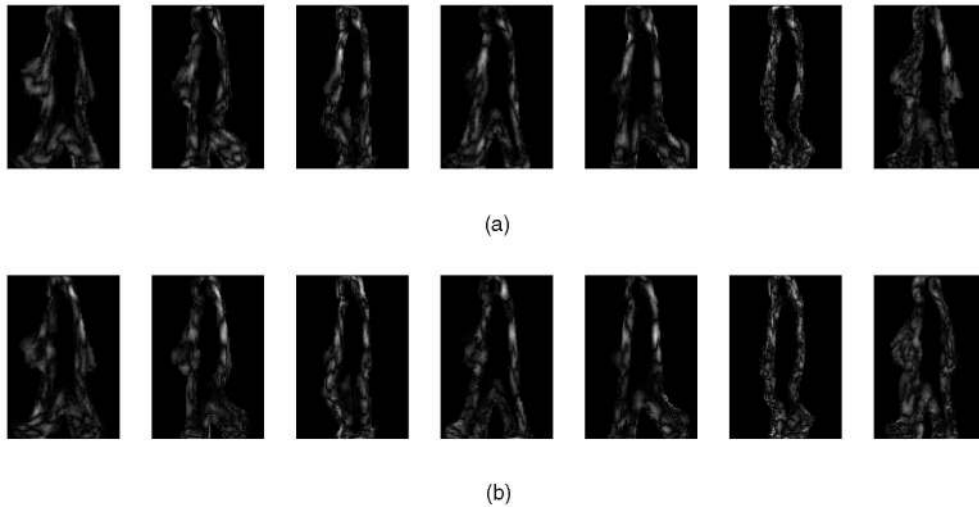


Fig 17. The top two most intersubject discriminating directions for each stance, as found by LDA of silhouette shapes from 33 subjects for that stance. Intrasubject variations span 16 combinations of variations in view, carrying condition, time, and surface.

dynamics normalization by itself only slightly improves performance over the baseline, when coupled with LDA the performance improvement is significant. From Fig. 16b, we see that the morphological operation-based distance computation also helps improve the recognition across surface type changes (experiments D, E, F, G and L), but slightly degrades the performance on other experiments.

The efficacy of dynamics normalization suggests that body-stance shape plays a more important role than dynamics in gait recognition. Note that we do use dynamics to time normalize the sequences, but shape plays a more significant role. This conclusion is also supported by Veeraraghavan et al. [19] who found that “kinematics does aid the task of recognition but pure kinematics is not enough for identification of an individual.” Another work that corroborates our observation is by Tolliver and Collins [15], who adopted a purely shape-based approach, relying on clustering stance shapes, with excellent results. To get some insight into the kinds of shape features that seem to be important, we consider the top two most intersubject discriminating directions for each stance, as found by LDA of the silhouette shapes in the training set used for gait-normalization. Fig. 17 shows these directions as images for some of the stances, with brightness proportional to the absolute value of the corresponding eigenvector components. Bright pixels are the important ones. We see 1) upper body, 2) knee, and 3) lower leg during gait swing phases seem to be the important features that are being picked up. For further improvements in gait recognition, it would be necessary to model and correct for the stance-shape changes under varying conditions. An excellent start is the work of Tanawongsuwan and Bobick [32], who studied the effect of walking speed on silhouette shapes.

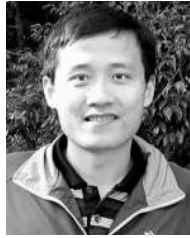
ACKNOWLEDGMENTS

This research was supported, in part, by funds from the US Defense Advanced Research Projects Agency HumanID at a Distance program (F49620-00-1-00388). The authors would like to thank the reviewers for their thoughtful comments and suggestions.

REFERENCES

- [1] J.E. Cutting and L.T. Kozlowski, “Recognition of Friends by Their Walk,” *Bull. of the Psychonomic Soc.*, vol. 9, pp. 353-356, 1977.
- [2] S. Niyogi and E. Adelson, “Analyzing Gait with Spatiotemporal Surfaces,” *Proc. IEEE Workshop Non-Rigid Motion*, pp. 24-29, 1994.
- [3] J.E. Boyd, “Synchronization of Oscillations for Machine Perception of Gaits,” *Computer Vision and Image Understanding*, vol. 96, pp. 35-59, Oct. 2004.
- [4] M.S. Nixon and J.N. Carter, “Advances in Automatic Gait Recognition,” *Proc. Int’l Conf. Automatic Face and Gesture Recognition*, pp. 139-146, 2004.
- [5] R. Gross and J. Shi, “The CMU Motion of Body (MoBo) Database,” Technical Report CMU-RI-TR-01-18, Robotics Inst., Carnegie Mellon Univ., June 2001.
- [6] A. Kale, A. Sundaresan, A.N. Rajagopalan, N.P. Cuntoor, A.K. Roy-Chowdhury, V. Krüger, and R. Chellappa, “Identification of Humans Using Gait,” *IEEE Trans. Image Processing*, vol. 13, pp. 1163-1173, Sept. 2004.
- [7] D.K. Wagg and M.S. Nixon, “On Automated Model-Based Extraction and Analysis of Gait,” *Proc. Int’l Conf. Automatic Face and Gesture Recognition*, pp. 11-16, 2004.
- [8] S. Sarkar, P.J. Phillips, Z. Liu, I. Robledo-Vega, P. Grother, and K.W. Bowyer, “The Human ID Gait Challenge Problem: Data Sets, Performance, and Analysis,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 2, pp. 162-177, Feb. 2005.
- [9] Z. Liu and S. Sarkar, “Effect of Silhouette Quality on Hard Problems in Gait Recognition,” *IEEE Trans. Systems, Man, and Cybernetics, Part B*, vol. 35, pp. 170-183, Apr. 2005.
- [10] Z. Liu, L. Malave, A. Osuntogun, P. Sudhakar, and S. Sarkar, “Toward Understanding the Limits of Gait Recognition,” *SPIE Proc. Defense and Security Symp.: Biometric Technology for Human Identification*, pp. 195-205, 2004.
- [11] L. Wang, T. Tan, H. Ning, and W. Hu, “Silhouette Analysis-Based Gait Recognition for Human Identification,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1505-1518, Dec. 2003.
- [12] S.D. Mowbry and M.S. Nixon, “Automatic Gait Recognition via Fourier Descriptors of Deformable Objects,” *Proc. Conf. Audio Visual Biometric Person Authentication*, pp. 566-573, 2003.
- [13] L. Lee, G. Dalley, and K. Tieu, “Learning Pedestrian Models for Silhouette Refinement,” *Proc. Int’l Conf. Computer Vision*, pp. 663-670, 2003.
- [14] R. Collins, R. Gross, and J. Shi, “Silhouette-Based Human Identification from Body Shape and Gait,” *Proc. Int’l Conf. Automatic Face and Gesture Recognition*, pp. 366-371, 2002.
- [15] D. Tolliver and R. Collins, “Gait Shape Estimation for Identification,” *Proc. Int’l Conf. Audio and Video-Based Biometric Person Authentication*, pp. 734-742, 2003.

- [16] A. Johnson and A. Bobick, "A Multi-View Method for Gait Recognition Using Static Body Parameters," *Proc. Int'l Conf. Audio and Video-Based Biometric Person Authentication*, pp. 301-311, 2001.
- [17] R. Tanawongsuwan and A. Bobick, "Performance Analysis of Time-Distance Gait Parameters Under Different Speeds," *Proc. Int'l Conf. Audio and Video-Based Biometric Person Authentication*, pp. 715-724, 2003.
- [18] J. Han and B. Bhanu, "Statistical Feature Fusion for Gait-Based Human Recognition," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 2, pp. 842-847, June 2004.
- [19] A. Veeraraghavan, A.R. Chowdhury, and R. Chellappa, "Matching Shape Sequences in Video with Applications in Human Movement Analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 12, pp. 1896-1909, Dec. 2005.
- [20] L. Lee and W. Grimson, "Gait Analysis for Recognition and Classification," *Proc. Int'l Conf. Automatic Face and Gesture Recognition*, pp. 155-162, 2002.
- [21] C.-S. Lee and A. Elgammal, "Gait Style and Gait Content: Bilinear Models for Gait Recognition Using Gait Re-Sampling," *Proc. Int'l Conf. Automatic Face and Gesture Recognition*, pp. 147-152, May 2004.
- [22] Z. Liu and S. Sarkar, "Simplest Representation Yet for Gait Recognition: Averaged Silhouette," *Proc. Int'l Conf. Pattern Recognition*, vol. 4, pp. 211-214, 2004.
- [23] Y. Liu, R. Collins, and Y. Tsin, "Gait Sequence Analysis Using Frieze Patterns," *Proc. European Conf. Computer Vision*, pp. 657-671, May 2002.
- [24] G. Zhao, R. Chen, G. Liu, and H. Li, "Amplitude Spectrum-Based Gait Recognition," *Proc. Int'l Conf. Automatic Face and Gesture Recognition*, pp. 23-28, 2004.
- [25] P.J. Phillips, H. Moon, S. Rizvi, and P. Rauss, "The FERET Evaluation Methodology for Face-Recognition Algorithms," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090-1104, Oct. 2000.
- [26] A. Sunderesan, A.K. RoyChowdhury, and R. Chellappa, "A Hidden Markov Model Based Framework for Recognition of Humans from Gait Sequences," *Proc. IEEE Int'l Conf. Image Processing*, vol. 2, pp. 93-96, 2003.
- [27] L. Rabiner and B.H. Juang, *Fundamentals of Speech Recognition*. Prentice Hall, 1993.
- [28] H. Akaike, "Information Theory as an Extension of the Maximum Likelihood Principle," *Proc. Second Int'l Symp. Information Theory*, pp. 267-281, 1973.
- [29] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *Proc. IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, July 1997.
- [30] L. Lee, "Gait Analysis for Classification," PhD thesis, Massachusetts Inst. of Technology, June 2003.
- [31] A. Kale, C.B.B. Yegnanarayana, A.N. Rajagopalan, and R. Chellappa, "Gait Analysis for Human Identification," *Proc. Int'l Conf. Audio and Video-Based Biometric Person Authentication*, pp. 706-714, 2003.
- [32] R. Tanawongsuwan and A. Bobick, "Modelling the Effects of Walking Speed on Appearance-Based Gait Recognition," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 783-790, June 2004.



Zongyi Liu received the BS degree in business from Shenzhen University in 1997, the MS degree in computer science and application from the University of Electronic Science and Technology of China in 2000, and the PhD degree in computer science from the University of South Florida in 2004, with research interests in computer-vision-based gait biometrics, pattern recognition, motion, and image segmentation. Currently, he is a software engineer at Amazon.com.



Sudeep Sarkar received the BTech degree in electrical engineering from the Indian Institute of Technology, Kanpur, in 1988. He received the MS and PhD degrees in electrical engineering, on a University Presidential Fellowship, from the Ohio State University, Columbus, in 1990 and 1993, respectively. Since 1993, he has been with the Computer Science and Engineering Department at the University of South Florida, Tampa, where he is currently a professor. His research interests include perceptual organization in single images and multiple image sequences, biometrics, gait recognition, color-texture analysis, and performance evaluation of vision systems. He has coauthored one book and coedited another book on perceptual organization. He is the recipient of the US National Science Foundation CAREER award in 1994, the USF Teaching Incentive Program Award for undergraduate teaching excellence in 1997, the Outstanding Undergraduate Teaching Award in 1998, and the Theodore and Vnette Askounes-Ashford Distinguished Scholar Award in 2004. He served on the editorial boards for the *IEEE Transactions on Pattern Analysis and Machine Intelligence* (1999-2003) and *Pattern Analysis & Applications Journal* during (2000-2001). He currently serves on the editorial board of the *Pattern Recognition* and the *IEEE Transactions on Systems, Man, and Cybernetics, Part-B*. He is a senior member of the IEEE.

► For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.