Individual Recognition Using Gait Energy Image

Ju Han, *Student Member*, *IEEE*, and
Bir Bhanu, *Fellow*, *IEEE*

Abstract—In this paper, we propose a new spatio-temporal gait representation, called Gait Energy Image (GEI), to characterize human walking properties for individual recognition by gait. To address the problem of the lack of training templates, we also propose a novel approach for human recognition by combining statistical gait features from real and synthetic templates. We directly compute the real templates from training silhouette sequences, while we generate the synthetic templates from training sequences by simulating silhouette distortion. We use a statistical approach for learning effective features from real and synthetic templates. We compare the proposed GEI-based gait recognition approach with other gait recognition approaches on USF HumanID Database. Experimental results show that the proposed GEI is an effective and efficient gait representation for individual recognition, and the proposed approach achieves highly competitive performance with respect to the published gait recognition approaches.

Index Terms—Gait recognition, real and synthetic templates, distortion analysis, feature fusion, performance evaluation, video.

1 Introduction

CURRENT image-based individual human recognition methods, such as fingerprints, face, or iris biometric modalities, generally require a cooperative subject, views from certain aspects, and physical contact or close proximity. These methods cannot reliably recognize noncooperating individuals at a distance in the real world under changing environmental conditions. Gait, which concerns recognizing individuals by the way they walk, is a relatively new biometric without these disadvantages. However, gait also has some limitations: it can be affected by clothing, shoes, or environmental context. Moreover, special physical conditions such as injury can also change a person's walking style. The large gait variation of the same person under different conditions (intentionally or unintentionally) reduces the discriminating power of gait as a biometric and it may not be as unique as fingerprint or iris, but the inherent gait characteristic of an individual [1] still makes it irreplaceable and useful in visual surveillance.

2 Related Work and Our Contribution

In recent years, various techniques have been proposed for human recognition by gait. These techniques can be divided as model-based and model-free approaches. In this paper, we focus on model-free approaches that do not recover a structural model of human motion. Little and Boyd [2] describe the shape of the human motion with scale-independent features from moments of the dense optical flow, and recognize individuals by phase vectors estimated from the feature sequences. Sundaresan et al. [3] proposed a hidden Markov models (HMMs) based framework for individual recognition by gait. Huang et al. [4] extend the template matching method to gait recognition by combining transformation based on canonical analysis and eigenspace transformation for feature selection. Sarkar et al. [5] directly measure the similarity between the gallery sequence and the probe sequence by computing the correlation of corresponding time-normalized frame pairs. Collins et al. [6] first extract key frames from a sequence and then the similarity between two sequences is computed using the normalized correlation.

In comparison with state-of-the-art, the contribution of this paper are:

- *New gait representation*—We propose a new spatio-temporal gait representation, called Gait Energy Image (GEI), for individual recognition [7]. Unlike other gait representations [4], [8] which consider gait as a sequence of templates, GEI represents human motion in a single image while preserving temporal information. In comparison to the gait representation by binary silhouette sequence, GEI not only saves storage space and computation time, but it is also less sensitive to silhouette noise in individual frames.
- *Synthetic templates*—To address the problem of lack of gallery gait data, we propose a simple but novel approach to generate synthetic GEI templates through silhouette distortion analysis. The synthetic GEI templates so-obtained are relatively insensitive to lower silhouette part distortion and small silhouette scale changes.
- *Feature fusion*—Individual recognition is performed by combining statistical gait features from real and synthetic templates [9]. The fused features not only characterize human walking properties under similar environmental conditions, but also predict gait properties under other conditions.
- *Experimental Results*—The proposed approach is tested on USF HumanID database silhouette version 1.7 and 2.1. We have provided the recognition performance based on real gait templates, synthetic gait templates, and fusion of real and synthetic templates, as well as the comparison with state-of-the-art published results.

3 Gait Energy Image (GEI) Representation

In this paper, we only consider individual recognition by activity-specific human motion, i.e., regular human walking, which is used in most current approaches of individual recognition by gait.

3.1 Motivation

Regular human walking can be considered as cyclic motion where human motion repeats at a stable frequency. While some gait recognition approaches [4] extract features from the correlation of all the frames in a walking sequence without considering their order, other approaches extract features from each frame and compose a feature sequence for the human walking sequence [2], [5], [6]. During the recognition procedure, these approaches either match the statistics collected from the feature sequence, or match the features between the corresponding pairs of frames in two sequences that are time-normalized with respect to their cycle lengths. The fundamental assumptions made here are: 1) the order of poses in human walking cycles is the same, i.e., limbs move forward and backward in a similar way among normal people, and 2) differences exist in the phase of poses in a walking cycle, the extend of limbs, and the shape of the torso, etc. Under these assumptions, it is possible to represent the spatio-temporal information in a single 2D gait template instead of an ordered image sequence.

3.2 Representation Construction

We assume that silhouettes have been extracted from original human walking sequences. A silhouette preprocessing procedure [5] is then applied on the extracted silhouette sequences. It includes size normalization (proportionally resizing each silhouette image so that all silhouettes have the same height) and horizontal alignment (centering the upper half silhouette part with respect to its horizontal centroid). In a preprocessed silhouette sequence, the time series signal of lower half silhouette size from each frame

- *The authors are with the Center for Research in Intelligent Systems, University of California, Riverside, 900 University Avenue, Riverside, CA 92521. E-mail: {jhan, bhanu}@cris.ucr.edu.*

Manuscript received 4 Oct. 2004; revised 9 June 2005; accepted 19 July 2005; published online 13 Dec. 2005.
Recommended for acceptance by S. Baker.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-0530-1004.

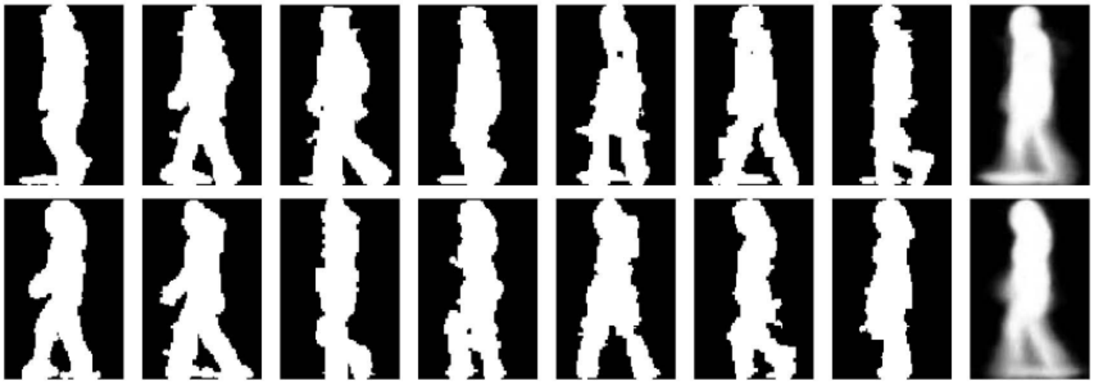

Fig. 1. Examples of normalized and aligned silhouette frames in different human walking sequences. The rightmost image in each row is the corresponding gait energy image (GEI).

indicates the gait frequency and phase information. We estimate the gait frequency and phase by maximum entropy spectrum estimation [2] from the time series signal.

Given the preprocessed binary gait silhouette images $B_t(x,y)$ at time t in a sequence, the gray-level gait energy image (GEI) is defined as follows:

$$G(x,y) = \frac{1}{N}\sum_{t=1}^{N} B_t(x,y), \tag{1}$$

where N is the number of frames in the complete cycle(s) of a silhouette sequence, t is the frame number in the sequence (moment of time), and x and y are values in the 2D image coordinate. Fig. 1 shows the sample silhouette images in a gait cycle from two people and the right most image is the corresponding GEI. As expected, GEI reflects major shapes of silhouettes and their changes over the gait cycle. We refer to it as gait energy image because: 1) each silhouette image is the space-normalized energy image of human walking at this moment, 2) GEI is the time-normalized accumulative energy image of human walking in the complete cycle(s), and 3) a pixel with higher intensity value in GEI means that human walking occurs more frequently at this position (i.e., with higher energy).

Bobick and Davis [8] propose motion-energy image (MEI) and motion-history image (MHI) for human movement type representation and recognition. Both MEI and MHI are vector-images where the vector value at each pixel is a function of the motion properties at this location in an image sequence. As compared to MEI and MHI, GEI targets specific normal human walking representation and we use GEI as the gait template for individual recognition.

3.3 Representation Justification

In comparison with the gait representation by binary silhouette sequence, GEI representation saves both storage space and computation time for recognition and is less sensitive to silhouette noise in individual frames. Consider a noisy silhouette image $B_t(x,y)$ that is formed by the addition of noise $\eta_t(x,y)$ to an original silhouette image $f_t(x,y)$, that is, $B_t(x,y) = f_t(x,y) + \eta_t(x,y)$, where we assume that at every pair of coordinates (x,y) the noise at different moments t is uncorrelated and identically distributed. Under these constraints, we further assume that $\eta_t(x,y)$ satisfies the following distribution:

$$\eta_t(x,y) = \begin{cases} \eta_{1t}(x,y) : P\{\eta_t(x,y) = -1\} = p, & \\ \quad P\{\eta_t(x,y) = 0\} = 1-p, & \text{if } f_t(x,y) = 1 \\ \eta_{2t}(x,y) : P\{\eta_t(x,y) = 1\} = p, & \\ \quad P\{\eta_t(x,y) = 0\} = 1-p, & \text{if } f_t(x,y) = 0. \end{cases} \tag{2}$$

We have

$$E\{\eta_t(x,y)\} = \begin{cases} -p, & \text{if } f_t(x,y) = 1 \\ p, & \text{if } f_t(x,y) = 0 \end{cases} \tag{3}$$

and

$$\sigma^2_{\eta_t(x,y)} = \sigma^2_{\eta_{1t}(x,y)} = \sigma^2_{\eta_{2t}(x,y)} = p(1-p). \tag{4}$$

Given a walking cycle with N frames where $f_t(x,y) = 1$ at a pixel (x,y) only in M frames, we have

$$\begin{aligned} G(x,y) &= \frac{1}{N}\sum_{t=1}^{N} B_t(x,y) \\ &= \frac{1}{N}\sum_{t=1}^{N} f_t(x,y) + \frac{1}{N}\sum_{t=1}^{N} \eta_t(x,y) = \frac{M}{N} + \bar{\eta}(x,y). \end{aligned} \tag{5}$$

Therefore, the noise in GEI is

$$\bar{\eta}(x,y) = \frac{1}{N}\sum_{t=1}^{N} \eta_t(x,y) = \frac{1}{N}\left[\sum_{t=1}^{M} \eta_{1t}(x,y) + \sum_{t=M+1}^{N} \eta_{2t}(x,y)\right]. \tag{6}$$

We have

$$\begin{aligned} E\{\bar{\eta}(x,y)\} &= \frac{1}{N}\left[\sum_{t=1}^{M} E\{\eta_{1t}(x,y)\} + \sum_{t=M+1}^{N} E\{\eta_{2t}(x,y)\}\right] \\ &= \frac{1}{N}[M(-p) + (N-M)p] = \frac{N-2M}{N}p \end{aligned} \tag{7}$$

and

$$\begin{aligned} \sigma^2_{\bar{\eta}(x,y)} &= E\{[\bar{\eta}(x,y) - E\{\bar{\eta}(x,y)\}]^2\} \\ &= \frac{1}{N^2} E\left\{\left[\sum_{t=1}^{M}[\eta_{1t}(x,y) - E\{\eta_{1t}(x,y)\}]\right.\right. \\ &\quad \left.\left. + \sum_{t=M+1}^{N}[\eta_{2t}(x,y) - E\{\eta_{2t}(x,y)\}]\right]^2\right\} \\ &= \frac{1}{N^2}\left[M\sigma^2_{\eta_{1t}(x,y)} + (N-M)\sigma^2_{\eta_{2t}(x,y)}\right] = \frac{1}{N}\sigma^2_{\eta_t(x,y)}. \end{aligned}$$

Therefore, the mean of the noise in GEI varies between $-p$ and p depending on M while its variability ($\sigma^2_{\bar{\eta}(x,y)}$) decreases. If $M = N$ at (x,y) (all $f_t(x,y) = 1$), $E\{\bar{\eta}(x,y)\}$ becomes $-p$; if $M = 0$ at (x,y) (all $f_t(x,y) = 0$), $E\{\bar{\eta}(x,y)\}$ becomes p. At the location (x,y), the mean of the noise in GEI is the same as that in the individual silhouette image, but the noise variance reduces so that the probability of outliers is reduced. If M varies between 0 and N at (x,y), $E\{\bar{\eta}(x,y)\}$ also varies between p and $-p$. Therefore, both the mean and the variance of the noise in GEI are reduced compared to the individual silhouette image at these locations. At the extreme, the noise in GEI has zero mean and reduced variance where $M = N/2$. As a result, GEI is less sensitive to silhouette noise in individual frames.

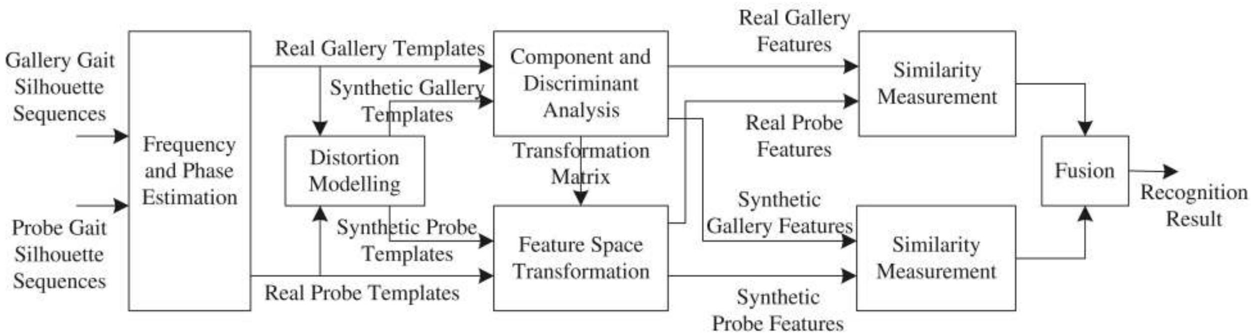

Fig. 2. System diagram of human recognition using the proposed statistical feature fusion approach.

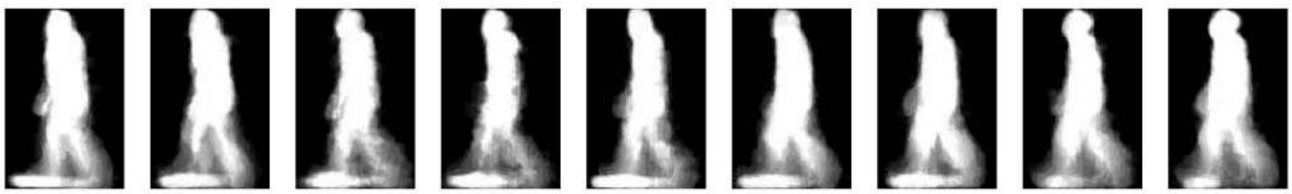

Fig. 3. An example of real gait template set generated from a long silhouette sequence of an individual.

4 Human Recognition Using GEI Templates

In this section, we describe the proposed statistical feature fusion approach for gait-based human recognition. In the training procedure, each gallery silhouette sequence is divided into cycles by frequency and phase estimation. Real gait templates are then computed from each cycle and distorted to generate synthetic gait templates. Next, we perform a component and discriminant analysis procedure on real and synthetic gait templates, respectively. As a result, real and synthetic transformation matrices and real features and synthetic features that form feature databases are obtained. In the recognition procedure, each probe silhouette sequence is processed to generate real and synthetic gait templates. These templates are then transformed by real and synthetic transformation matrices to obtain real and synthetic features, respectively. Probe features are compared with gallery features in the database, and a feature fusion strategy is applied to combine real and synthetic features at the decision level to improve recognition performance. The system diagram is shown in Fig. 2.

4.1 Real and Synthetic Gait Templates

The number of training sequences for each person is limited (one or several) in real surveillance applications. This makes it difficult to recognize individuals under various other conditions not exhibited in the data. To solve this problem, one solution is to directly measure the similarity between the gallery (training) and probe (testing) templates. However, direct template matching is sensitive to silhouette distortions such as scale and displacement changes. Statistical feature learning may extract inherent properties of training templates from an individual and, therefore, it will be less sensitive to such silhouette distortion. However, with gait templates obtained under similar conditions, the learned features may overfit the training data. Therefore, to overcome these problems, we generate two sets of gait templates—real templates and synthetic templates.

The real gait templates for an individual are directly computed from each cycle of the silhouette sequence of this individual. Let $\{R_i\}$, $i = 1, \ldots, n_R$, be the real GEI template set of the individual, where n_R is the number of completes cycles in the silhouette sequence. Fig. 3 shows an example of the real GEI template set from a long gait sequence of an individual. Note the similarity of template appearance in the presence of noise.

Although real gait templates provide cues for individual recognition, all the templates from the same sequence are obtained under the "same" physical conditions. If the conditions change, the learned features may not work well for recognition. Various conditions affect the silhouette appearance from the same person: walking surface, shoe, clothing, etc. The common silhouette distortion in the lower part of the silhouette occurs under most conditions. This kind of distortion includes shadows, missing body parts, and sequential silhouette scale changes. For example, silhouettes on the grass surface may miss the bottom part of feet, while silhouettes on the concrete surface may contain strong shadows. In these cases, silhouette size normalization errors occur, and silhouettes so-obtained may have different scales with respect to silhouettes on other surfaces. Therefore, we generate a series of synthetic gait templates (see Fig. 5) that are less sensitive to the distortion in the lower silhouette and small silhouette scale changes. Note that a moving object detection approach can also provide information about the material type on which a person is walking [10].

Let $R_0 = \frac{1}{n_R}\sum_{i=1}^{n_R} R_i$ be the fundamental GEI template computed from n_R cycles of a given silhouette sequence (the leftmost image in Fig. 5). Assume that k bottom rows of R_0 are missed due to some kind of environmental conditions. According to the silhouette preprocessing procedure in Section 3.2, the remaining part needs to be proportionally resized to fit to the original height. In the same way, we can generate a series of new synthetic GEI templates corresponding to different lower body part distortion.

Synthetic gait templates are computed from R_0 of a given silhouette sequence by following a distortion model based on anthropometric data [11]. The length from the bottom of bare foot to the ankle above the sole is approximately $1/24$ of the stature. Considering the height of heelpiece and shadow, we select $2/24$ of the silhouette height from the bottom of an original GEI template and the associated width as an estimate of the allowable distortion for all original training GEI templates. For all original testing GEI templates, we use $3/24$ of the silhouette height for distortion. We allow larger distortion for testing templates since we want to allow larger distortion in unknown situations. The pseudocode for generating synthetic GEI templates is shown in Fig. 4.

The synthetic templates ($\{S_i\}$, $i = 1, \ldots, n_S$ shown in Fig. 5) expanded from the same R_0 have similar global shape properties but different bottom parts and different scales. Therefore, they can be effectively used for individual recognition in the presence of silhouette scale changes and lower silhouette distortion encountered in the real-world applications.

1. Given an original GEI template of size $X \times Y$
2. Let h be the highest row from the bottom corresponding to the maximum allowable distortion
3. Let $k = 2$
4. Initialize $i = 1$
5. Remove $r = k * i$ rows from the bottom of the original template
6. Resize the remaining template from $(X - r) \times Y$ to $X \times \frac{XY}{X-r}$ by nearest neighbor interpolation
7. Equally cut left and right borders to generate a synthetic template S_i of size $X \times Y$
8. Let $i = i + 1$
9. If $k * i \leq h$, go to step 5; otherwise, stop

Fig. 4. Pseudocode for generating synthetic GEI templates.

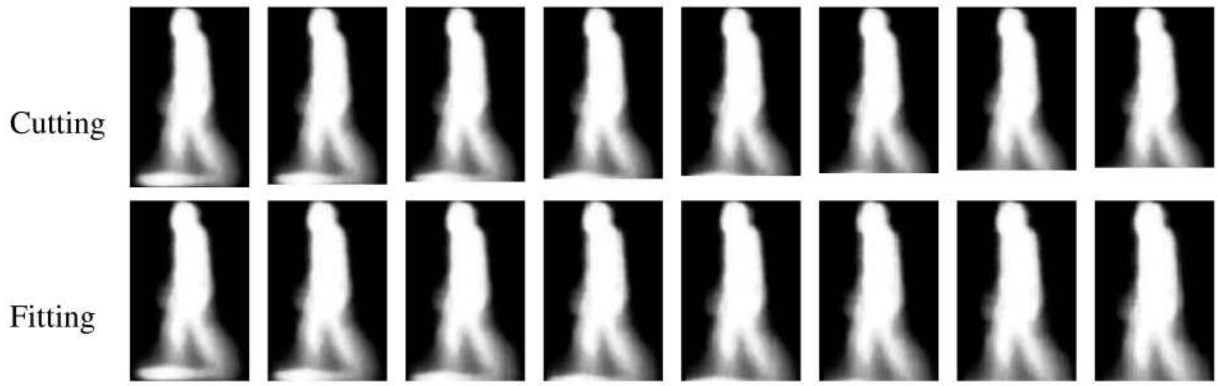

Fig. 5. Procedure of generating synthetic GEI templates from an original template. The leftmost template is the original template, and the other templates are generated by gradually cutting the bottom portion (templates in the first row) and fitting it to the original template size (templates in the second row). Synthetic templates in the second row are used as the synthetically generated gait templates.

In order to find effective features, we use a statistical feature extraction method to learn gait features from real and synthetic templates. Features learned from real templates characterize human walking properties provided in training sequences and features learned from synthetic templates simulate gait properties under other real-world conditions that incur distortion in the lower (feet) part of the silhouette.

4.2 Learning Gait Features by Component and Discriminant Analysis

Once we obtain a series of training GEI templates (real or synthetic) for each individual, the problem of their excessive dimensionality occurs. There are two classical linear approaches for finding transformations for dimensionality reduction—Principal Component Analysis (PCA) and Multiple Discriminant Analysis (MDA) that have been effectively used in face recognition [12]. PCA seeks a projection that best represents the data in the least-square sense, while MDA seeks a projection that best separates the data in the least-square sense. Huang et al. [4] combine PCA and MDA to achieve the best data representation and the best class separability simultaneously. In this paper, the learning procedure follows this combination approach.

Given n d-dimensional training GEI templates $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$, PCA minimizes the function

$$J_{d'} = \sum_{k=1}^{n} \left\| \left(\mathbf{m} + \sum_{i=1}^{d'} a_{ki} \mathbf{e}_i \right) - \mathbf{x}_k \right\|^2, \tag{8}$$

where $d' < d$, $\mathbf{m} = \frac{1}{n}\sum_{k=1}^{n} \mathbf{x}_k$, and $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{d'}\}$ are a set of orthogonal unit vectors. $J_{d'}$ is minimized when $\mathbf{e}_1, \mathbf{e}_2, \ldots,$ and $\mathbf{e}_{d'}$ are the d' eigenvectors of the scatter matrix $S = \sum_{i=1}^{n}(\mathbf{x}_i - \mathbf{m})(\mathbf{x}_i - \mathbf{m})^T$ having the largest eigenvalues. The d'-dimensional feature vector $\mathbf{y}_k$ is obtained from the $\mathbf{x}_k$ as follows:

$$\mathrm{y}_k = M_{pca}\mathrm{x}_k = [a_1, \ldots, a_{d'}]^T = [\mathbf{e}_1, \ldots, \mathbf{e}_{d'}]^T \mathrm{x}_k, \quad k = 1, \ldots, n. \tag{9}$$

Suppose that the n d'-dimensional principal component vectors $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n\}$ belong to c classes. MDA seeks a transformation matrix W that maximizes the ratio of the between-class scatter matrix S_B to the within-class scatter matrix S_W:

$$J(W) = \frac{|\tilde{S}_B|}{|\tilde{S}_W|} = \frac{|W^T S_B W|}{|W^T S_W W|}. \tag{10}$$

The within-class scatter matrix S_W is defined as $S_W = \sum_{i=1}^{c} S_i$, where $S_i = \sum_{\mathbf{y}\in\mathcal{D}_i}(\mathbf{y} - \mathbf{m}_i)(\mathbf{y} - \mathbf{m}_i)^T$ and $\mathbf{m}_i = \frac{1}{n_i}\sum_{\mathbf{y}\in\mathcal{D}_i}\mathbf{y}$, where $\mathcal{D}_i$ is the training template set that belongs to the ith class and n_i is the number of templates in $\mathcal{D}_i$. The between-class scatter S_B is defined as $S_B = \sum_{i=1}^{c} n_i(\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T$, where $\mathbf{m} = \frac{1}{n}\sum_{\mathbf{y}\in\mathcal{D}}\mathbf{y}$. $J(W)$ is maximized when the columns of W are the generalized eigenvectors that correspond to the largest eigenvalues in

$$S_B \mathbf{w}_i = \lambda_i S_W \mathbf{w}_i. \tag{11}$$

There are no more than $c - 1$ nonzero eigenvalues and the corresponding eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_{c-1}$ form a transformation matrix. The $(c-1)$-dimensional feature vector $\mathbf{z}_k$ is obtained from the d'-dimensional principal component vector $\mathbf{y}_k$:

$$\mathrm{z}_k = M_{mda}\mathrm{y}_k = [\mathbf{v}_1, \ldots, \mathbf{v}_{c-1}]^T \mathrm{y}_k, \quad k = 1, \ldots, n. \tag{12}$$

For each training gait template, its gait feature vector is obtained as follows:

$$\mathrm{z}_k = M_{mda}M_{pca}\mathrm{x}_k = T\mathrm{x}_k, \quad k = 1, \ldots, n. \tag{13}$$

The obtained feature vectors represent the n templates for individual recognition.

4.3 Individual Recognition

We train the real gait templates and synthetic gait templates separately for feature extraction. Let $\{\mathbf{r}\}$ be the set of real feature vectors extracted from real training gait templates, and T_r be the corresponding real transformation matrix. Similarly, let $\{\mathbf{s}\}$ be the set of synthetic feature vectors extracted from synthetic training gait templates, and T_s is the synthetic transformation matrix. The class centers for $\{\mathbf{r}\}$ and $\{\mathbf{s}\}$ are $\mathrm{m}_{ri} = \frac{1}{n_i}\sum_{\mathbf{r}\in\mathcal{R}_i}\mathbf{r}$ and $\mathrm{m}_{si} = \frac{1}{m_i}\sum_{\mathbf{s}\in\mathcal{S}_i}\mathbf{s}$,

TABLE 1
Twelve Experiments Designed for Individual Recognition in USF HumanID Database

Experiment Label	A	B	C	D	E	F	G	H	I	J	K	L
Size of Probe Set	122	54	54	121	60	121	60	120	60	120	33	33
Gallery/Probe Difference	V	H	VH	S	SH	SV	SHV	B	BS	BV	THC	TS

(Legends: V-View, H-Shoe, S-Surface, B-Briefcase, T-Time, and C-Clothing)

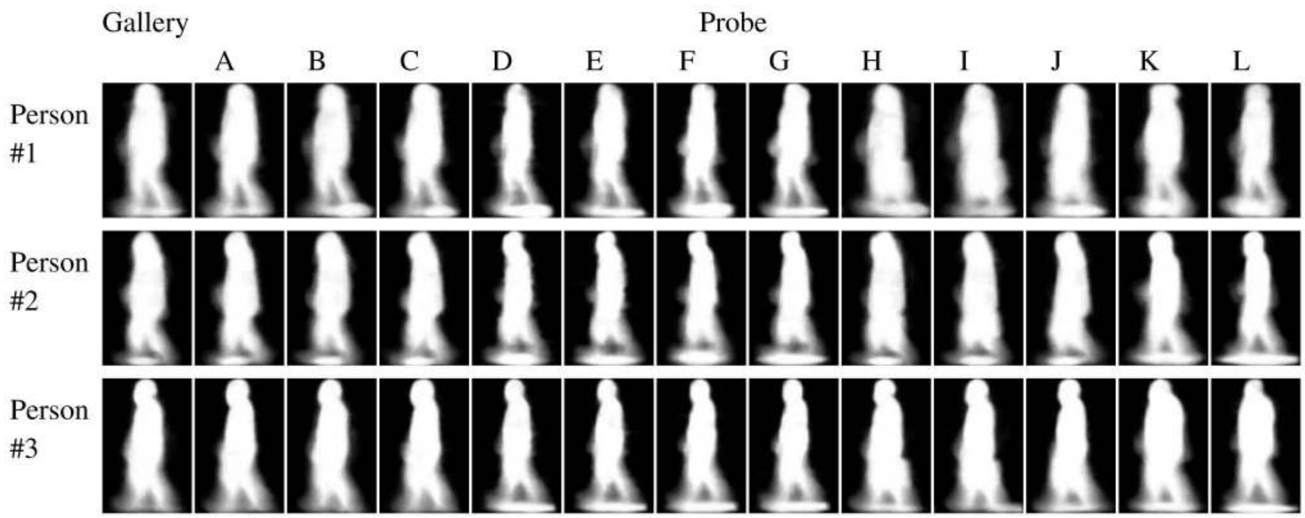

Fig. 6. GEI examples in USF HumanID database.

where $i = 1, \ldots, c$, c is the number of classes (individuals) in the database, $\mathcal{R}_i$ is the set of real feature vectors belonging to the ith class, $\mathcal{S}_i$ is the set of synthetic feature vectors belonging to the ith class, n_i is the number of feature vectors in $\mathcal{R}_i$, and m_i is the number of feature vectors in $\mathcal{S}_i$. Assuming that feature vectors in each class are Gaussian distributed with the same covariance matrix $\Sigma = \sigma^2 I$, Bayesian classifier becomes minimum Euclidean distance classifier that is used for individual recognition.

Given a probe gait silhouette sequence P, we follow the procedure in Section 4.1 to generate real gait templates $\{R_j\}$, $j = 1, \ldots, n_R$ and synthetic gait templates $\{S_j\}$, $j = 1, \ldots, n_S$. The corresponding real and synthetic feature vector sets are obtained as follows:

$$\begin{aligned} \{\hat{\mathcal{R}}_P\} &: \quad \hat{\mathrm{r}}_j = T_r R_j, \quad j = 1, \ldots, n_R \text{ and} \\ \{\hat{\mathcal{S}}_P\} &: \quad \hat{\mathrm{s}}_j = T_s S_j, \quad j = 1, \ldots, n_S. \end{aligned} \tag{14}$$

For the classifier based on real gait templates, we define

$$D(\hat{\mathcal{R}}_P, \mathcal{R}_i) = \frac{1}{n_R} \sum_{j=1}^{n_R} ||\hat{\mathrm{r}}_j - \mathrm{m}_{ri}||, \quad i = 1, \ldots, c. \tag{15}$$

We assign $P \in \omega_k$ if

$$D(\hat{\mathcal{R}}_P, \mathcal{R}_k) = \min_{i=1}^{c} D(\hat{\mathcal{R}}_P, \mathcal{R}_i). \tag{16}$$

For the classifier based on synthetic gait templates, we define

$$D(\hat{\mathcal{S}}_P, \mathcal{S}_i) = \min_{j=1}^{n_S} ||\hat{\mathrm{s}}_j - \mathrm{m}_{si}||, \quad i = 1, \ldots, c. \tag{17}$$

We assign $P \in \omega_k$ if

$$D(\hat{\mathcal{S}}_P, \mathcal{S}_k) = \min_{i=1}^{c} D(\hat{\mathcal{S}}_P, \mathcal{S}_i). \tag{18}$$

For the fused classifier, we define

$$\begin{aligned} &D(\{\hat{\mathcal{R}}_P, \hat{\mathcal{S}}_P\}, \{\mathcal{R}_i, \mathcal{S}_i\}) = \\ &\frac{c(c-1) D(\hat{\mathcal{R}}_P, \mathcal{R}_i)}{2 \sum_{i=1}^{c} \sum_{j=1, j \neq i}^{c} D(\mathcal{R}_i, \mathcal{R}_j)} + \frac{c(c-1) D(\hat{\mathcal{S}}_P, \mathcal{S}_i)}{2 \sum_{i=1}^{c} \sum_{j=1, j \neq i}^{c} D(\mathcal{S}_i, \mathcal{S}_j)}, \quad i = 1, \ldots, c, \end{aligned} \tag{19}$$

where $2 \sum_{i=1}^{c} \sum_{j=1, j \neq i}^{c} D(\mathcal{R}_i, \mathcal{R}_j)/c(c-1)$ is the average distance between real feature vectors of every two classes in the database which is used to normalize $D(\hat{\mathcal{R}}_P, \mathcal{R}_i)$ and $2 \sum_{i=1}^{c} \sum_{j=1, j \neq i}^{c} D(\mathcal{S}_i, \mathcal{S}_j)/c(c-1)$ has the similar meaning for the synthetic features. We assign $P \in \omega_k$ if

$$D(\{\hat{\mathcal{R}}_P, \hat{\mathcal{S}}_P\}, \{\mathcal{R}_k, \mathcal{S}_k\}) = \min_{i=1}^{c} D(\{\hat{\mathcal{R}}_P, \hat{\mathcal{S}}_P\}, \{\mathcal{R}_i, \mathcal{S}_i\}). \tag{20}$$

5 Experimental Results

5.1 Data and Parameters

Our experiments are carried out on the USF HumanID gait database [5]. This database consists of persons walking in elliptical paths in front of the camera. For each person, there are up to five covariates: viewpoints (left/right), two different shoe types, surface types (grass/concrete), carrying conditions (with/without a briefcase), and time and clothing. Twelve experiments are designed for individual recognition as shown in Table 1. The gallery set contains 122 sequences/individuals. Individuals are unique in the gallery and each probe set and there are no common sequences between the gallery set and any of the probe sets. Also, all the probe sets are distinct. The real GEIs (R_0 as mentioned in Section 4.1) of three individuals in the gallery set and their corresponding sequences in probe sets A-L are shown in Fig. 6.

Sarkar et al. [5] propose a baseline approach to extract human silhouette and recognize an individual in this database. For comparison, they provide extracted silhouette data which can be found at http://marathon.csee.usf.edu/GaitBaseline/. Our experiments begin with these extracted binary silhouette data. Silhouette data of version 1.7 (extracted from the parameterized algorithm) contains the gallery set (71 sequences/individuals) and probe set A-G only, while silhouette

TABLE 2
Comparison of Recognition Performance on Silhouette Sequence Version 2.1

	Rank1 Performance				Rank5 Performance			
		This paper				This paper		
	baseline	real	synthetic	fusion	baseline	real	synthetic	fusion
A	73%	89%	84%	90%	88%	93%	93%	94%
B	78%	87%	93%	91%	93%	93%	96%	94%
C	48%	78%	67%	81%	78%	89%	93%	93%
D	32%	36%	53%	56%	66%	65%	75%	78%
E	22%	38%	45%	64%	55%	60%	71%	81%
F	17%	20%	30%	25%	42%	42%	54%	56%
G	17%	28%	34%	36%	38%	45%	53%	53%
H	61%	62%	48%	64%	85%	88%	78%	90%
I	57%	59%	57%	60%	78%	79%	82%	83%
J	36%	59%	39%	60%	62%	80%	64%	82%
K	3%	3%	21%	6%	12%	6%	33%	27%
L	3%	6%	24%	15%	15%	9%	42%	21%

(Legends: baseline—direct frame shape matching [5]; real—proposed real gait feature classifier only; synthetic—proposed synthetic gait feature classifier only; fusion—proposed gait feature fusion.)

TABLE 3
Comparison of Recognition Performance Using Different Approaches on Silhouette Sequence Version 1.7

	Rank1 Performance						Rank5 Performance					
				This paper						This paper		
	USF	CMU	UMD	real	sync	fusion	USF	CMU	UMD	real	sync	fusion
A	79%	87%	99%	100%	97%	100%	96%	100%	100%	100%	99%	100%
B	66%	81%	89%	85%	88%	90%	80%	90%	92%	85%	90%	93%
C	56%	66%	78%	80%	80%	85%	76%	83%	92%	88%	88%	93%
D	29%	21%	36%	30%	42%	47%	61%	59%	62%	55%	70%	79%
E	24%	19%	29%	33%	48%	57%	52%	50%	54%	55%	69%	69%
F	30%	27%	24%	21%	33%	32%	45%	53%	47%	41%	58%	70%
G	10%	23%	18%	29%	40%	31%	33%	43%	48%	48%	66%	67%

(Legends: USF—direct frame shape matching [5]; CMU—key frame shape matching [6]; UMD—HMM framework [3]; real—proposed real gait feature classifier only; sync—proposed synthetic gait feature classifier only; fusion—proposed gait feature fusion.)

data of version 2.1 (extracted from the parameter-free algorithm) contains the gallery set (122 sequences/individuals) and all probe sets. The results on data from version 1.7 and 2.1 are shown in Tables 2 and 3, respectively.

There are two parameters in our proposed approach: the size of distortion area for generating synthetic templates and the number of principal components d' in (8). The former parameter has been discussed in Section 4.1. The latter parameter d' is chosen to facilitate the solution of (11). If $d' < c$, where c is the number of classes, or d' is too large (some elements in principal component vectors are too small), the S_W matrix becomes uninvertable. We choose $d' = 2c$ in our approach.

5.2 Performance Evaluation

The experimental results as well as comparison with other approaches of individual recognition by gait are shown in Tables 2 and 3. In these tables, rank1 performance means the percentage of the correct subjects appearing in the first place of the retrieved rank list and rank5 means the percentage of the correct subjects appearing in any of the first five places of the retrieved rank list. The performance in these tables is the recognition rate under these two definitions.

We perform experiments using 1) real features (obtained GEI without distortion), 2) synthetic features, and 3) fused features according to rules in (16), (18), and (20), respectively. Table 2 compares the recognition performance of USF baseline algorithm [5] and our proposed approach. It can be seen that the rank1 performance of proposed real feature classifier is better than or equivalent to that of baseline algorithm on all experiments. The rank5 performance of real feature classifier is better than that of the baseline algorithm on most experiments but slightly worse on D. This shows the inherent representational power of GEI and demonstrates that matching features learned from real gait templates achieve better recognition performance than direct matching between individual silhouette frame pairs in the baseline algorithm.

The performance of proposed synthetic feature classifier is significantly better than that of real feature classifier on experiments D-G and K-L. Probe sets in D-G have the common difference of walking surface with respect to the gallery set and probe sets in K-L have the common difference of time with respect to the gallery set. In these probe sets, there is silhouette distortion in the lower body part compared with silhouettes in the gallery set. As expected, the experimental results show that the proposed synthetic feature classifier is insensitive to this kind of distortion compared with the real feature classifier. However, the proposed synthetic feature classifier sacrifices the performance on experiments H-J where probe sets contain people who are carrying briefcases (as compared to the gallery set). The distortions as a result of briefcase occur beyond the selected distortion area in this paper.

The fused feature classifier achieves better performance than individual real feature classifier and synthetic feature classifier in most experiments, and achieves significantly better performance (both rank1 and rank5) than the baseline algorithm in all experiments. This shows that the fusion approach is effective and takes the advantage of merits in individual features.

Although the proposed fusion approach achieves significantly better results than the baseline algorithm, its performance is still not satisfactory in the presence of large silhouette distortion such as probe sets K and L. Examining the columns K and L in Fig. 6, note that K and L are quite different from the gallery in time, shoe and clothing, and time and surface, respectively. This requires a more complex model and analysis for distortion in these cases.

Table 3 compares the recognition performance of different published approaches on silhouette version 1.7. The rank1 and rank5 performance of real feature classifier is better than other approaches in A, C (rank1 only), E, and G, and slightly worse in B, D, and F. The rank1 and rank5 performance of synthetic feature classifier is better than other approaches in almost all the experiments but slightly worse than UMD HMM approach in A and B. The proposed fusion approach takes advantage of real and synthetic features and, therefore, achieves better performance (both rank1 and rank5) than other approaches in all the experiments.

6 CONCLUSIONS

In this paper, we propose a new spatio-temporal gait representation, called the Gait Energy Image (GEI), for individual recognition by gait. Unlike other gait representations which consider gait as a sequence of templates (poses), GEI represents human motion sequence in a single image while preserving temporal information. To overcome the limitation of training templates, we propose a simple model for simulating distortion in synthetic templates and a statistical gait feature fusion approach for human recognition by gait. Experimental results show that 1) GEI is an effective and efficient gait representation and 2) the proposed recognition approach achieves highly competitive performance with respect to the published major gait recognition approaches.

This paper presents a systematic and comprehensive gait recognition approach, which can work just as fine as other complex published techniques in terms of effectiveness of performance while providing all the advantages associated with the computational efficiency for real-world applications. Therefore, we believe that our technique will have an impact on practical applications.

ACKNOWLEDGMENTS

This work was supported in part by grants F49620-02-1-0315 and DAAD19-01-0357; the contents and information do not necessarily reflect the position or policy of the US government. The authors would like to thank P. Jonathon Phillips (NIST), Patrick Grother (NIST), and Sudeep Sarkar (USF) for their help in providing them with the HumanID data set used in this paper.

REFERENCES

[1] J. Han and B. Bhanu, “Performance Prediction for Individual Recognition by Gait,” *Pattern Recognition Letters*, vol. 26, no. 5, pp. 615-624, Apr. 2005.

[2] J.J. Little and J.E. Boyd, “Recognizing People by Their Gait: The Shape of Motion,” *Videre: J. Computer Vision Research*, vol. 1, no. 2, pp. 1-32, 1998.

[3] A. Sundaresan, A. Roy-Chowdhury, and R. Chellappa, “A Hidden Markov Model Based Framework for Recognition of Humans from Gait Sequences,” *Proc. Int'l Conf. Image Processing*, vol. 2, pp. 93-96, 2003.

[4] P.S. Huang, C.J. Harris, and M.S. Nixon, “Recognizing Humans by Gait via Parameteric Canonical Space,” *Artificial Intelligence in Eng.*, vol. 13, pp. 359-366, 1999.

[5] S. Sarkar, P.J. Phillips, Z. Liu, I.R. Vega, P. Grother, and K.W. Bowyer, “The Humanid Gait Challenge Problem: Data Sets, Performance, and Analysis,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 2, pp. 162-177, Feb. 2005.

[6] R.T. Collins, R. Gross, and J. Shi, “Silhouette-Based Human Identification from Body Shape and Gait,” *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition*, pp. 351-356, 2002.

[7] J. Han and B. Bhanu, “Gait Energy Image Representation: Comparative Performance Evaluation on USF HumanID Database,” *Proc. Joint Int'l Workshop VS-PETS*, pp. 133-140, Nov. 2003.

[8] A.F. Bobick and J.W. Davis, “The Recognition of Human Movement Using Temporal Templates,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 3, pp. 257-267, Mar. 2001.

[9] 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, 2004.

[10] S. Nadimi and B. Bhanu, “Physical Models for Moving Shadow and Object Detection in Video,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 26, no. 8, pp. 1079-1087, Aug. 2004.

[11] S. Pheasant, *Bodyspace: Anthropometry, Ergonomics, and Design.* London and Philadelphia: Taylor & Francis, 1986.

[12] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, “Eigenfaces versus Fisherfaces: Recognition Using Class Specific Linear Projection,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, July 1997.