Individual Recognition Using Gait Energy Image

Ju Han and Bir Bhanu
Center for Research in Intelligent Systems
University of California, Riverside, California 92521, USA
{jhan,bhanu}@cris.ucr.edu

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 generate a series of new GEI templates by analyzing the human silhouette distortion under various conditions. Principal component analysis followed by multiple discriminant analysis are used for learning features from the expanded GEI training templates. Recognition is carried out based on the learned features. 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 current gait recognition approaches.

1. Introduction

Current human recognition methods, such as fingerprints, face or iris biometrics, generally require a cooperative subject, views from certain aspects and physical contact or close proximity. These methods can not reliably recognize non-cooperating individuals at a distance in real-world changing environmental conditions. Moreover, in various applications of personal identification, many established biometrics can be obscured. Gait, which concerns recognizing individuals by the way they walk, has been an important biometric without the above-mentioned disadvantages.

In this paper, we propose a new spatio-temporal gait representation, Gait Energy Image (GEI), for individual recognition. Unlike other gait representations [8, 4] which consider gait as a sequence of templates (poses), GEI represents human motion sequence in a single image while preserving some temporal information. We also propose a statistical approach to learn and recognize individual gait properties from the limited training GEI templates.

In the next section, we introduce related work of human recognition by gait. The representation of GEI is introduced in Section 3. In Section 4, we propose two approaches for human recognition using GEI: direct GEI matching, and statistical GEI feature matching. In Section 5, we analyze the experimental results of the proposed human recognition approaches and compare them with the existing techniques. Section 6 concludes the paper.

2. Related Work

In recent years, various approaches have been proposed for human recognition by gait. These approaches can be divided into two categories: model-based approaches and model-free approaches.

2.1. Model-based Approaches

When people observe human walking patterns, they not only observe the global motion properties, but also interpret the structure of the human body and detect the motion patterns of local body parts. The structure of the human body is generally interpreted based on their prior knowledge. Model-based gait recognition approaches focus on recovering a structural model of human motion, and the gait patterns are then generated from the model parameters for recognition.

Niyogi and Adelson [14] make an initial attempt in a spatiotemporal (XYT) volume. They first find the bounding contours of the walker, and then fit a simplified stick model on them. A characteristic gait pattern in XYT is generated from the model parameters for recognition. Yoo et al. [19] estimate hip and knee angles from the body contour by linear regression analysis. Then trigonometric-polynomial interpolant functions are fitted to the angle sequences, and the parameters so-obtained are used for recognition. In Lee and Grimson's work [11], human silhouette is divided into local regions corresponding to different human body parts, and ellipses are fitted to each region to represent the human structure. Spatial and spectral features are extracted from these local regions for recognition and classification. Bhanu and Han [3] propose a kinematic-based approach to recognize individuals by gait. The 3D human walking parameters are estimated by performing a least squares fit of the 3D kinematic model to the 2D silhouette extracted from a monocular image sequence. Human gait signatures are generated by selecting features from the estimated parameters.

In these model-based approaches, the accuracy of human model reconstruction strongly depends on the quality of the extracted human silhouette. In the presence of noise, the estimated parameters may not be reliable. To obtain more reliable estimates, Tanawongsuwan and Bobick [17] reconstruct the human structure by tracking 3D sensors attached on fixed joint positions. However, their approach needs lots of human interaction which is not applicable in most surveillance applications.

2.2. Model-free Approaches

Model-free approaches make no attempt to recover a structural model of human motion. The features used for gait representation includes: moments of shape, height and stride/width, and other image/shape templates.

Moments of shape is one of the most commonly used gait features. Little and Boyd [12] describe the shape of human motion with a set of features derived from moments of a dense flow distribution. Shutler et al. [16] include velocity into the traditional moments to obtain the so-called velocity moments (VMs). A human motion image sequence can be represented as a single VM value with respect to a specific moment order instead of a sequence of traditional moment values for each frame. He and Debrunner's [7] approach detects a sequence of feature vectors based on Hu's moments of each motion segmented frame, and the individual is recognized from the feature vector sequence using hidden Markov models (HMMs).

BenAbdelkader et al. [2] use height, stride and cadence as features for human identification. Kale et al. [10] choose the width vector from the extracted silhouette as the representation of gait. Continuous HMMs are trained for each person and then used for gait recognition. In their later work [9], different gait features are further derived from the width vector and recognition is performed by a direct matching algorithm.

To avoid the feature extraction process which may reduce the reliability, Murase and Sakai [13] propose a template matching method to calculate the spatio-temporal correlation in a parametric eigenspace representation for gait recognition. Huang et al. [8] extend this approach by combining transformation based on canonical analysis, with eigenspace transformation for feature selection. BenAbdelkader et al. [1] compute the self-similarity plot by correlating each pair of aligned and scaled human silhouette in an image sequence. Normalized features are then generated from the similarity plots and used for gait recognition via eigenspace transformation.

As a direct template matching approach, Phillips et al. [15] measure the similarity between the gallery sequence and the probe sequence by computing the correlation of corresponding time-normalized frame pairs. Similarly, Collins et al. [5] first extract key frames from a sequence, and the similarity between two sequences is computed from normalized correlation. Tolliver and Collins [18] cluster human silhouettes/poses of each training sequence into k prototypical shapes. In the recognition procedure, the silhouettes in a testing sequence are also classified into k prototypical shapes which are compared to prototypical shapes of each training sequence for similarity measurement.

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. Some gait recognition approaches extract features from the correlation of all the frames in a walking sequence without considering their order [1, 8, 13]. Other approaches extract features from each frame and compose a feature sequence for the human walking sequence [2, 5, 7, 10, 9, 12, 15, 16, 18]. During the recognition procedure, they either match the extracted statistics 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, respectively. The assumption here is that the order of poses in human walking cycles is the same, i.e., the limbs (arms and legs) move forward and backward in a similar way among normal people. The difference exists in the phase of poses in a walking cycle, the extend of limbs, and the shape of the torso, etc. As the order of poses in regular human walking is generally not considered in gait recognition approaches, it is possible to compose a spatio-temporal template in a single image instead of a ordered image sequences as usual.

3.2. Representation Construction

We use a silhouette extraction procedure and begin with the extracted binary silhouette sequences. The preprocess-

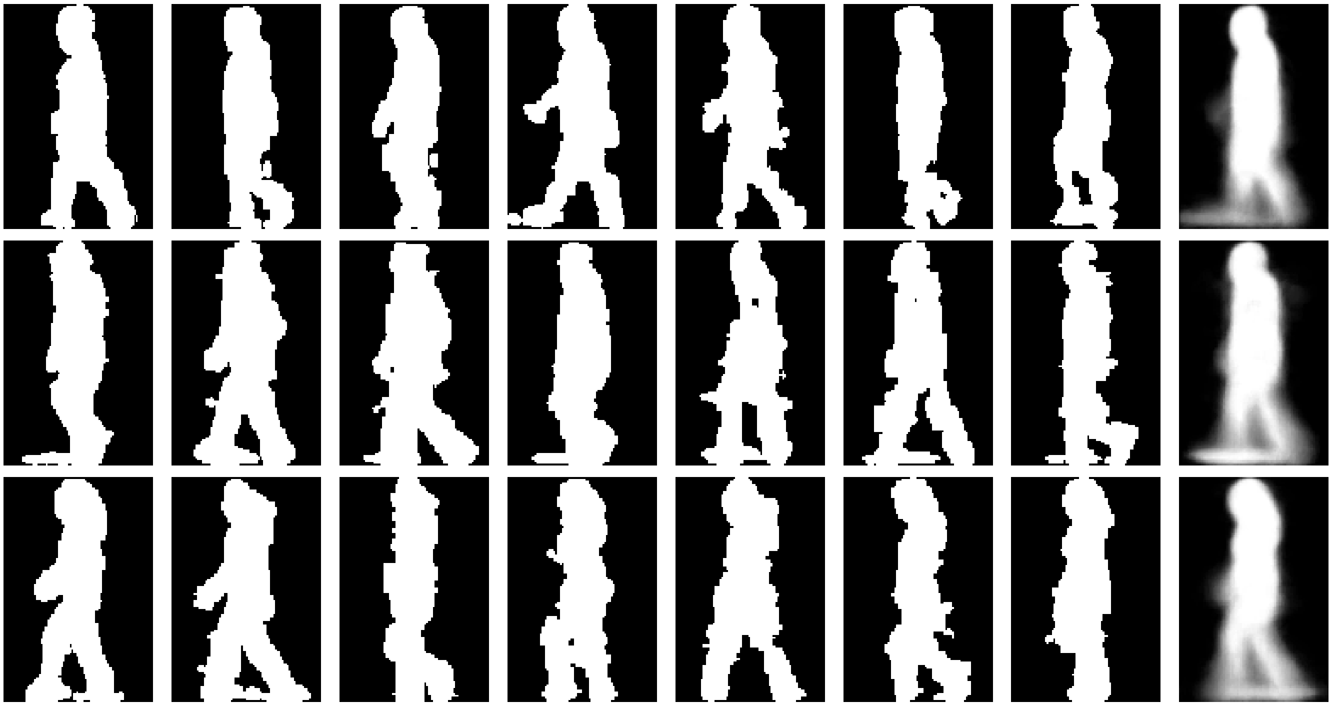

Figure 1. Examples of normalized and aligned silhouette frames in different human walking sequences. The rightmost image in each row is the average silhouette image over the whole sequence - Gait Energy Image (GEI).

ing procedure includes size normalization – fitting the silhouette height to the fixed image height, and sequential horizontal alignment – centering the upper half silhouette part with respect to the horizontal centroid. Figure 1 shows examples of preprocessed silhouette frames in different human walking sequences. The rightmost image in each row is the average silhouette image over the whole sequence. As expected, the average silhouette image reflects the major shapes of the human silhouettes and their changes over the sequence. A pixel with higher intensity value means that human body occurs more frequently at this position. Therefore, we refer to this average silhouette image as Gait Energy Image (GEI).

Given a size-normalized and horizontal-aligned human walking binary silhouette sequence $B(x, y, t)$, the grey-level GEI $G(x, y)$ is defined as follows

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

where N is the number of frames in complete cycles of the sequence, t is the frame number of the sequence, x and y are values in the 2D image coordinate.

3.3. Representation Justification

GEI has several advantages over the representation of binary silhouette sequence. As an average template, GEI is not sensitive to incidental silhouette errors in individual frames. The robustness could be further improved if we discard those pixels with the energy values lower than a threshold. Moreover, with such a 2D template, we do not need to divide the silhouette sequence into cycles and perform time normalization with respect to the cycle length. Therefore, the errors occuring in these procedures can be therefore avoid.

Compared with binary silhouette sequence, the information loss of GEI is obvious. For a specific pixel in GEI, we only know its intensity value, i.e., the frequency with which the human silhouette occurs at this position over the whole sequence. However, we might partly reconstruct the original silhouette sequence from the GEI according to the knowledge of regular human walking. For example, for a pixel near the outline of the leg area, it GEI value shows that silhouette occurs at this location in 20 frames out of 100 frames. Using the common sense, we know that 20 frames should be those frames where human stride instead of standing straight, if noise is not considered. Similarly, we can allocate the GEI values to most other leg/arm areas to corresponding frames in the silhouette sequence. In general, the energy changes in the torso and head area can be considered as noise. Although the knowledge is not enough to accurately allocate the GEI value of each pixel (i.e., the original silhouette sequence cannot be completely reconstructed), GEI still keeps the major shapes of human walking and reflects the major shape changes during walking. Actually, it is difficult to analyze how and in what degree the information loss affects the discriminating power of GEI as a template for individual recognition. We will evaluate this issue in the section of experimental results by comparing the recognition performance between GEI and binary silhouette sequence representations.

3.4. Relationship with MEI and MHI

Bobick and Davis [4] propose motion-energy image (MEI) and motion-history image (MHI) for human movement recognition. Both MEI and MHI are vector-image where the vector value at each pixel is a function of the motion properties at this location in an image sequence.

MEI is a binary image which represents where motion has occured in an image sequence:

$$E_\tau(x,y,t) = \cup_{i=0}^{\tau-1} D(x,y,t-i), \tag{2}$$

where $D(x,y,t)$ is a binary sequence indicating regions of motion, τ is the duration of time, t is the moment of time, x and y are values of 2D image coordinate. To represent a regular human walking sequence, if $D(x,y,t)$ is normalized and aligned as $B(x,y,t)$ in Equation (1), MEI $E_N(x,y,N)$ is the binary version of GEI $G(x,y)$.

MHI is a grey-level image which represents how motion in the image is moving:

$$H_\tau(x,y,t) = \begin{cases} \tau, & \text{if} \quad D(x,y,t) = 1; \\ \max\{0, H_\tau(x,y,t-1) - 1\}, & \text{otherwise.} \end{cases} \tag{3}$$

In general, both MEI and MHI are different motion representations compared to GEI. As regular human walking is a cyclic and highly self-occluded motion with a specific style, MEI and MHI are not suitable to represent regular human walking for individual recognition.

4. Human Recognition Using GEI Templates

Human walking sequences for training are limited in real surveillance applications. Because each sequence is represented as one GEI template, the training/gallery GEIs for each individual might limited to several or even one template(s). In this paper, we develop two approaches to recognize individuals from the limited templates.

4.1. Direct GEI Matching

One possible approach is recognizing individuals by measuring the similarity between the gallery (training) and probe (testing) templates. Given GEIs of two gait sequences, $G_g(x,y)$ and $G_p(x,y)$, their distance can be measured by calculating their normalized matching error:

$$D(G_g, G_p) = \frac{\sum_{x,y} |G_g(x,y) - G_p(x,y)|}{\sqrt{\sum_{x,y} G_g(x,y) \sum_{x,y} G_p(x,y)}}, \tag{4}$$

where $\sum_{x,y} |G_g(x,y) - G_p(x,y)|$ is the matching error between two GEIs, $\sum_{x,y} G_g(x,y)$ and $\sum_{x,y} G_p(x,y)$ are total energy in two GEIs, respectively.

This direct GEI matching approach is sensitive to distortion in silhouettes generated from image sequences that are recorded under different conditions. Recognition by learning may recover the inherent properties in training templates from an individual and therefore insensitive to such silhouette distortion. However, with one GEI template per individual, learning cannot be performed. Even with several templates per individual, if they are from similar conditions, the learned features may be overfit to the training templates.

4.2. Statistical GEI Feature Matching

In this section, we propose a statistical GEI feature matching approach for individual recognition from limited GEI templates. First, we generate new templates from the limited training templates according to a distortion analysis. Next, statistical features are learned from the expanded training templates by principal component analysis (PCA) to reduce the dimension of the template and multiple discriminant analysis (MDA) to achieve better class seperatability. The individual is recognized by the learned features. The system diagram of training and recognition procedure is shown in Figure 2.

4.2.1 Generating New Templates from Limited Training Templates

Various factors have effect on silhouettes extracted from the same person: shoe and clothing, walking surface, camera view, and shadow, etc. Shoe, surface and shadow affect the foot area of the silhouette. In addition, shoe and surface also change the human walking style. Clothing affects the shape of the silhouette. If the camera view changes slightly, there will be slight changes in silhouettes; if the camera view changes a lot, the extracted silhouettes may be totally different which may cause recognition to fail.

Among these factors, slight camera view changes may be neglected. The silhouette shape distortion incurred by the difference of clothing is irregular distortion, which occurs

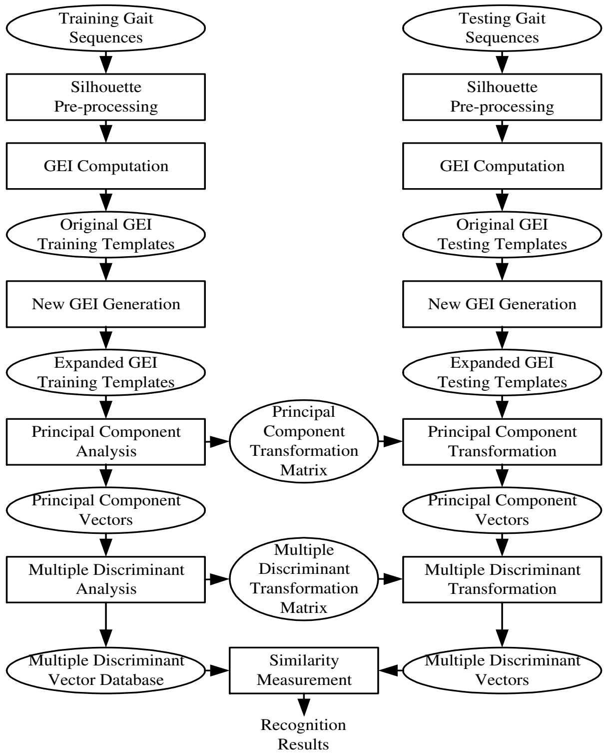

Figure 2. System diagram of individual recognition using the proposed statistical GEI feature matching approach.

in the upper body, lower body or both, and make body parts fatter or thinner. Thus, it is difficult to model this irregular distortion. Similarly, different shoes and walking surfaces incur global silhouette distortions which are also difficult to model. Now we consider the common distortion incurred by the difference of shoe, surface and shadow which generally occurs in the foot area of the silhouette. These distortions are local distortions which make the bottom part of the silhouette and GEI unreliable. If we generate new templates which are insensitive to the distortion in their bottom parts, the learned template properties will be insensitive to this kind of distortion.

The new GEI templates are generated as illustrated in Fig 3. First, we determine the range of the distortion area, e.g., n rows from the bottom row of the original GEI. Then, we cut a portion of the area from the bottom, and fit it to the original GEI size to obtain a new template. By repeating this step until reaching the upper row of the distortion area, we will obtain a series of new templates. The training templates expanded from the same original GEI have the same global shape properties but different bottom parts and different scales. Therefore, the learned features from the expanded training templates are insensitive to the common distortion by shadow, shoe and surface which occurs in the bottom part of GEI templates.

4.2.2 Learning Templates by Component Analysis and Discriminants

Once we obtain a series of training GEI templates for each individual, the problem of their excessive dimensionality occurs. To reduce their dimensionality, there are two classical approaches of finding effective linear transformations by combing features - Principal Component Analysis (PCA) and Multiple Discriminant Analysis (MDA). As described in [6], PCA seeks a projection that best represents the data in a least square sense, while MDA seeks a projection that best separates the data in a least-square sense. Huang et al. [8] 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 templates $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n\}$, PCA minimizes the criterion function

$$J_{d'} = \sum_{k=1}^{n} ||(\mathbf{m} + \sum_{i=1}^{d'} a_{ki}\mathbf{e}_i) - \mathbf{x}_k||^2, \tag{5}$$

where $d' < d$, $\mathbf{m} = \frac{1}{n}\sum_{k=1}^{n} \mathbf{x}_k$, and $\{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_{d'}\}$ are a set of unit vectors. $J_{d'}$ is minimized when $\mathbf{e}_1$, $\mathbf{e}_2$, ..., and $\mathbf{e}_{d'}$ are the d' eigenvectors of the scatter matrix S having the largest eigenvalues, where

$$S = \sum_{k=1}^{n} (\mathbf{x}_k - \mathbf{m})(\mathbf{x}_k - \mathbf{m})^T. \tag{6}$$

The d'-dimensional principal component vector $\mathbf{y}_k$ is obtained from the d-dimensional GEI template $\mathbf{x}_k$ by multiplying the transformation matrix $[\mathbf{e}_1, ..., \mathbf{e}_{d'}]$:

$$\mathrm{y}_k = [a_1, ..., a_{d'}]^T = [\mathbf{e}_1, ..., \mathbf{e}_{d'}]^T \mathrm{x}_k, \quad k = 1, ..., n. \tag{7}$$

where n is the number of the expanded GEI templates from all people in the training dataset.

Although PCA finds components that are useful for representing data, there is no reason to assume that these components must be useful for discriminating between data in different classes because PCA does not consider the class label of training templates. Multiple discriminant analysis (MDA) seeks a projection that are efficient for discrimination. Suppose that the n d'-dimensional transformed training templates $\{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n\}$ belong to c classes. MDA seeks a transformation matrix W that in some sense maximizes the ratio of the between-class scatter S_B to the within-class scatter S_W:

$$J(W) = \frac{|\tilde{S}_B|}{|\tilde{S}_W|} = \frac{|W^T S_B W|}{|W^T S_W W|}. \tag{8}$$

The within-class scatter S_B is defined as

$$S_W = \sum_{i=1}^{c} S_i, \tag{9}$$

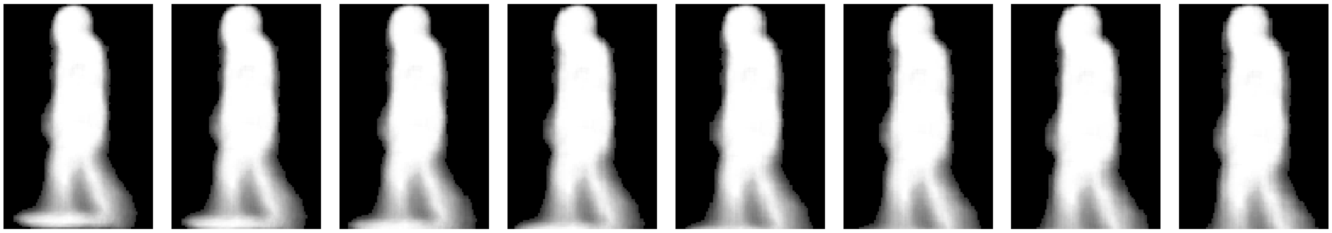

Figure 3. Examples of new GEI templates generated from the original template. The leftmost template is the original template, other templates are sequentially generated by cutting the bottom portion (2 rows in this example) of the previous template and fitting it to the original template size.

where

$$S_i = \sum_{\mathbf{y} \in \mathcal{D}_i} (\mathbf{y} - \mathbf{m}_i)(\mathbf{y} - \mathbf{m}_i)^T \quad (10)$$

and

$$\mathbf{m}_i = \frac{1}{n_i} \sum_{\mathbf{y} \in \mathcal{D}_i} \mathbf{y}, \quad (11)$$

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 within-class scatter S_B is defined as

$$S_B = \sum_{i=1}^{c} n_i (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T, \quad (12)$$

where

$$\mathbf{m} = \frac{1}{n} \sum_{\mathbf{y} \in \mathcal{D}} \mathbf{y}, \quad (13)$$

and $\mathcal{D}$ is the whole training template set. $J(W)$ is maximized when the columns of W are the generalized eigenvectors that correspond to the largest eigenvalues in

$$S_B \mathbf{w}_i = \lambda_i S_W \mathbf{w}_i. \quad (14)$$

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

$$\mathrm{z}_k = [\mathbf{v}_1, ..., \mathbf{v}_{c-1}]^T \mathrm{y}_k, \quad k = 1, ..., n. \quad (15)$$

The obtained multiple discriminant vectors compose the feature database for individual recognition.

4.2.3 Individual Recognition

Given the GEI template $\tilde{\mathbf{x}}$ of a query gait sequence, a set of n_q templates $\{\tilde{\mathbf{x}}_1, ..., \tilde{\mathbf{x}}_{n_q}\}$ are generated according to the procedure described in Section 4.2.1. After the principal component transformation and multiple discriminant transformation, we obtain a set of feature vectors $\{\tilde{\mathbf{z}}_1, ..., \tilde{\mathbf{z}}_{n_q}\}$ for this test gait sequence. The feature distance between the query gait sequence and each class in the feature database can be given by the minimum distance between query and training feature vector pairs as follows

$$Distance_i = \min_{\mathbf{z} \in \mathcal{D}_i} \min_{j=1}^{n_q} \sum_{k=1}^{c-1} |\tilde{z}_{jk} - z_k|, \quad i = 1, ..., c. \quad (16)$$

After the distances for all classes are obtained, they are ranked in an ascending order where the class with the smallest distance is the best match of the query gait sequence.

5. Experimental Results

Our experiments are carried out on the USF HumanID May-2001 gait database. This database consists of 452 sequences from 74 persons walking in elliptical paths in front of the cameras. For each person, there are up to 5 covariates: viewpoints - Left/Right, shoe types - A/B, surface types - grass/concrete, carrying conditions - with/without a briefcase, and time and clothing. Seven experiments are designed for individual recognition as shown in Table 1. The gallery set contains 71 sequences. No sequence belongs to the same person in each individual data set.

Phillips et al. [15] proposed a baseline approach to extract human silhouettes and recognize individuals in this database. For comparison, they provide extracted silhouette data which can be found at the website http://marathon.csee.usf.edu/GaitBaseline/. Our experiments begin with these extracted binary silhouette data (parameterized version 1.7). The experimental results are shown in Table 2 and 3 as well as comparison with other approaches of individual recognition by gait. In these tables, rank1 means that only the first subject in the retrieval rank list is recognized as the same subject as the query subject, and rank5 means that the first five subjects are all recognized as the same subject as the query subject. The performance in these tables is the recognition rate under these two definitions.

Table 1. Seven experiments designed for individual recognition in USF HumanID database.

Experiment Label	Size of Probe Set	Difference between Gallery and Probe Sets
A	71	View
B	41	Shoe
C	41	View and Shoe
D	70	Surface
E	44	Surface and Shoe
F	70	Surface and View
G	44	Surface, Shoe and View

Table 2. Comparison of recognition performance of Rank 1 among different approaches on silhouette sequence version 1.7. (Legends: USF - direct frame shape matching [15]; DGEI - direct GEI matching, this paper; CMU - key frame shape matching [5]; SPS1/SPS2 - clustered frame shape matching with two criteria [18]; SGEI - statistical GEI feature matching, this paper.)

	USF	DGEI	CMU	SPS1	SPS2	SGEI
A	79%	99%	87%	82%	85%	90%
B	66%	83%	81%	66%	81%	90%
C	56%	73%	66%	54%	60%	73%
D	29%	18%	21%	20%	23%	41%
E	24%	14%	19%	18%	17%	40%
F	30%	11%	27%	21%	25%	27%
G	10%	10%	23%	21%	21%	38%

5.1. Recognition Results by Direct GEI Matching

To evaluate the effectiveness of GEI as a gait representation, we carry out experiments of individual recognition by direct matching between GEI templates according to the distance metric give by Equation (4). As we mentioned in Section 2.2, Phillips et al. [15] measure the similarity between the gallery sequence and the probe sequence by computing the correlation of corresponding time-normalized frame pairs. This approach can be viewed as a typical direct matching approach between regular gait silhouette sequences. We compare the recognition performance between their approach (USF) and our direct GEI matching approach (DGEI) as shown in Table 2 and 3.

The left part of Table 2 and 3 shows the recognition performance of USF and DGEI approaches. It is shown that our DEGI approach achieves much better results in experiments A-C. In these experiments, the difference between gallery and probe data exists in view, shoe or both, which incur little distortion in extracted silhouette. This means that GEI is less sensitive to this kind of distortion than regular gait silhouette sequence.

Table 3. Comparison of recognition performance of Rank 5 among different approaches on silhouette sequence version 1.7. (Same legend as in Table 2)

	USF	DGEI	CMU	SPS1	SPS2	SGEI
A	96%	100%	100%	98%	90%	99%
B	80%	93%	90%	90%	87%	93%
C	76%	93%	83%	81%	80%	93%
D	61%	55%	59%	46%	52%	68%
E	52%	52%	50%	43%	43%	69%
F	45%	47%	53%	46%	48%	58%
G	33%	52%	43%	43%	44%	60%

Although the rank1 performance of DGEI and USF are both not good in experiments D-G, our DEGI is worse than that of USF (See Table 2). The probe sets in experiments D-G have the common difference of surface with respect to the gallery set. As we discussed previously, the distortion incurred by surface difference is relatively high. For example, if the same person walks at different surface, the extracted silhouettes may have different shadows. In addition, the silhouette from a walking sequence on the grass surface may miss the bottom part of the feet because they could be covered by the grass. In this case, silhouette height normalization errors occur, and the silhouette so-obtained may have different scale with respect to the silhouette on other surfaces. It is shown that the GEI is sensitive to this kind of distortion with respect to the regular silhouette sequence. However, the rank5 performance of our DGEI is similar to that of USF in experiments D-G (See Table 3). This shows that GEI is competitive with regular silhouette sequence because the rank1 results are not reliable and more ranked subjects should be considered in these experiments. Another reason of the rank1 worse performance of DGEI (See Table 3) is that silhouettes of version 1.7 are not well-aligned.

5.2. Recognition Results by Statistical GEI Feature Matching

Table 2 and 3 show that our individual recognition approach by statistical GEI feature matching (SGEI) achieves better recognition results than DGEI in the experiments with large silhouette distortion, i.e., D-G. In other experiments with small silhouette distortion, the performance of SGEI is better than that of DGEI in experiments B and C, but slightly worse in experiments A. Thus SGEI slightly sacrifices the performance in experiments with small silhouette distortion while improving the performance in experiments with large silhouette distortion with respect to DGEI.

We also compare the performance of SGEI with other approaches published in [15, 5, 18] in Table 2 and 3. It is shown that SGEI achieves better or equivalent recognition performance than other approaches in all experiments.

6. Conclusions

In this paper, a new spatio-temporal gait representation, called Gait Energy Image (GEI), is proposed for individual recognition by gait. Unlike other gait representation which considers gait as a sequence of templates (poses), GEI represents a human motion sequence in a single image while preserving temporal information. To overcome the limitation of training templates, we generate a series of new GEI templates by analyzing the human silhouette distortion under various conditions. Principal component analysis and multiple discriminant analysis are used for learning features from the expanded GEI training templates. Recognition is then carried out based on the learned features. Experimental results show that (a) GEI is an effective and efficient gait representation which is insensitive to incidental silhouette errors in individual frames, and (b) the proposed recognition approach achieves highly competitive performance with respect to the published gait recognition approaches.

Acknowledgment

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 U.S. Government. The authors would like to thank Dr. Sudeep Sarkar (University of Southern Florida) and Dr. Patrick Grother (National Institute of Standards and Technology) for their help in providing the HumanID dataset and silhouette data.

References

[1] C. BenAbdelkader, R. Cutler, and L. Davis. Motion-based recognition of people in eigengait space. in *Proc. International Conference on Automatic Face and Gesture Recognition*, pages 254–259, May 2002.

[2] C. BenAbdelkader, R. Cutler, and L. Davis. Person identification using automatic height and stride estimation. in *Proc. International Conference on Pattern Recognition*, 4:377–380, 2002.

[3] B. Bhanu and J. Han. Individual recognition by kinematic-based gait analysis. in *Proc. International Conference on Pattern Recognition*, 3:343–346, 2002.

[4] A. Bobick and J. Davis. The recognition of human movement using temporal templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(3):257–267, March 2001.

[5] R. Collins, R. Gross, and J. Shi. Silhouette-based human identification from body shape and gait. in *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, pages 351–356, May 2002.

[6] R. Duda, P. Hart, and D. Stork. *Pattern Classification*. John Willy & Sons, 2000.

[7] Q. He and C. Debrunner. Individual recognition from periodic activity using hidden markov models. in *Proc. IEEE Workshop on Human Motion*, pages 47–52, 2000.

[8] P. Huang, C. Harris, and M. Nixon. Recognizing humans by gait via parameteric canonical space. *Artificial Intelligence in Engineering*, 13:359–366, 1999.

[9] A. Kale, N. Cuntoor, B. Yegnanarayana, A. Rajagopalan, and R. Chellappa. Gait analysis for human identification. in *Proc. 4th International Conference on Audio- and Video-Based Biometric Person Authentication, LNCS 2688*, pages 706–714, 2003.

[10] A. Kale, A. Rajagopalan, N. Cuntoor, and V. Kruger. Gait-based recognition of humans using continuous hmms. in *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, pages 321–326, May 2002.

[11] L. Lee and W. Grimson. Gait analysis for recognition and classification. in *Proc. 5th International Conference on Automatic Face and Gesture Recognition*, pages 148–155, May 2002.

[12] J. Little and J. Boyd. Recognizing people by their gait: the shape of motion. *Videre: Journal of Computer Vision Research*, 1(2):1–32, 1998.

[13] H. Murase and R. Sakai. Moving object recognition in eigenspace representation: gait analysis and lip reading. *Pattern Recognition Letters*, 17(2):155–62, 1996.

[14] S. Niyogi and E. Adelson. Analyzing and recognizing walking figures in xyt. in *Proc. IEEE Conference on CVPR*, pages 469–474, 1994.

[15] P. Phillips, S. Sarkar, I. Robledo, P. Grother, and K. Bowyer. The gait identification challenge problem: data sets and baseline algorithm. in *Proc. IEEE International Conference on Pattern Recognition*, 1:385–388, 2002.

[16] J. Shutler, M. Nixon, and C. Harris. Statistical gait recognition via velocity moments. in *Proc. IEE Colloquium on Visual Biometrics*, pages 10/1–10/5, March 2000.

[17] R. Tanawongsuwan and A. Bobick. Gait recognition from time-normalized joint-angle trajectories in the walking plane. in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2:726–731, 2001.

[18] D. Tolliver and R. Collins. Gait shape estimation for identification. in *Proc. 4th International Conference on Audio- and Video-Based Biometric Person Authentication, LNCS 2688*, pages 734–742, 2003.

[19] J.-H. Yoo, M. Nixon, and C. Harris. Model-driven statistical analysis of human gait motion. in *Proc. IEEE International Conference on Image Processing*, 1:285–288, Sept. 2002.