Silhouette Analysis-Based Gait Recognition for Human Identification

Liang Wang, Tieniu Tan, *Senior Member*, *IEEE*, Huazhong Ning, and Weiming Hu

Abstract—Human identification at a distance has recently gained growing interest from computer vision researchers. Gait recognition aims essentially to address this problem by identifying people based on the way they walk. In this paper, a simple but efficient gait recognition algorithm using spatial-temporal silhouette analysis is proposed. For each image sequence, a background subtraction algorithm and a simple correspondence procedure are first used to segment and track the moving silhouettes of a walking figure. Then, eigenspace transformation based on *Principal Component Analysis* (PCA) is applied to time-varying distance signals derived from a sequence of silhouette images to reduce the dimensionality of the input feature space. Supervised pattern classification techniques are finally performed in the lower-dimensional eigenspace for recognition. This method implicitly captures the structural and transitional characteristics of gait. Extensive experimental results on outdoor image sequences demonstrate that the proposed algorithm has an encouraging recognition performance with relatively low computational cost.

Index Terms—Human motion analysis, biometrics, gait recognition, principal component analysis.

1 Introduction

Visual analysis of human motion [17], [25] attempts to detect, track, and identify people, and, more generally, to understand human behaviors from image sequences involving humans. Biometrics is a technology that makes use of the physiological or behavioral characteristics to authenticate the identities of people [18]. The combination of human motion analysis and biometrics in surveillance systems has become a popular research direction over the past few years. Vision-based human identification at a distance, in particular, has recently gained wider interest from the computer vision community. This interest is strongly driven by the need for automated person identification systems for visual surveillance and monitoring applications in security-sensitive environments such as banks, parking lots, and airports. An ongoing research project, the Human ID program [1] initialized by DARPA (*Defense Advanced Research Project Agency*), aims to develop a full range of multimode surveillance technologies for successfully detecting, classifying, and identifying humans to enhance the protection of facilities from terrorist attacks. Its focus is on dynamic face recognition and recognition from body dynamics including gait.

Different from gait classification which classifies human motion into categories, such as walking, running, and jumping, gait recognition, also called gait-based human identification, is a relatively new research direction in biometrics. It aims to discriminate individuals by the way they walk. In comparison with other first-generation biometric features such as fingerprint and iris, gait has the advantage of being unobtrusive, i.e., it requires no subject contact other than walking, in common with automatic face recognition [15]. So far, gait is probably the only perceivable biometric feature from a great distance. Each person seems to have a distinctive and idiosyncratic way of walking, which can be easily understood from a biomechanics viewpoint [20], [21]. Human ambulation consists of synchronized integrated movements of hundreds of muscles and joints [20]. Although these basic patterns of bipedal locomotion are similar for humans, gaits do vary from one person to another in certain details such as their relative timing and magnitudes [20]. There have been some allied studies such as biomechanics [21], physical medicine studies for therapy [38], psychological studies [37], and approaches aiming to model and track a human through an image sequence, though usually not for recognition [17], [25], [2], [15]. All these related subjects lend strong support to the potential for gait as a useful cue of biometrics.

1.1 Overview of Approach

Our study aims to establish an automatic gait recognition method based upon spatiotemporal silhouette analysis measured during walking. Gait includes both the body appearance and the dynamics of human walking motion [34]. Intuitively, recognizing people by gait depends greatly on how the silhouette shape of an individual changes over time in an image sequence. So, we may consider gait motion to be composed of a sequence of static body poses and expect that some distinguishable signatures with respect to those static body poses can be extracted and used for recognition by considering temporal variations of those observations. Also, eigenspace transformation based on PCA has actually been demonstrated to be a potent metric in face recognition (i.e., eigenface) and gait analysis [6], [7], [11], [31]. Based on these observations, this paper proposes a silhouette analysis-based gait recognition algorithm using the traditional PCA. The algorithm implicitly captures the structural and transitional characteristics of gait. Although

• *The authors are with the National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences, Beijing, P.R. China, 100080. E-mail: {lwang, tnt, hzning, wmhu}@nlpr.ia.ac.cn.*

Manuscript received 31 July 2003; revised 1 Feb. 2003; accepted 23 June 2003. Recommended for acceptance by S. Sarkar.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 117040.

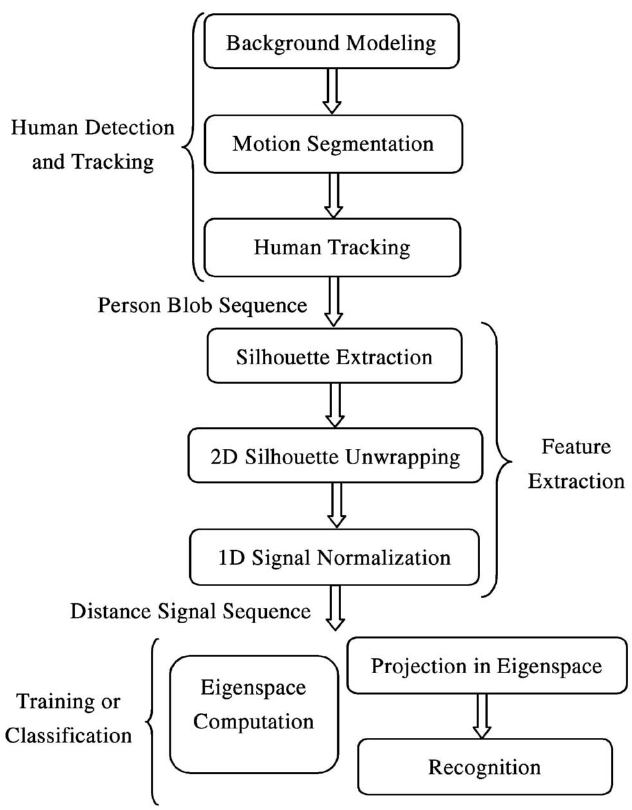


Fig. 1. Overview of the proposed method.

it is very simple in essence, the experimental results are surprisingly promising.

The overview of the proposed algorithm is shown in Fig. 1. It consists of three major modules, namely, human detection and tracking, feature extraction, and training or classification. The first module serves to detect and track the walking figure in an image sequence. A background subtraction procedure is performed to segment motion from the background, and the moving region corresponding to the spatial silhouette of the walking figure is successively tracked through a simple correspondence method. The second module is used to extract the binary silhouette from each frame and map the 2D silhouette image into a 1D normalized distance signal by contour unwrapping with respect to the silhouette centroid. Accordingly, the shape changes of these silhouettes over time are transformed into a sequence of 1D distance signals to approximate temporal changes of gait pattern. The third module either applies PCA on those time-varying distance signals to compute the predominant components of gait signatures (training phase), or determines the person's identity using the standard nonparametric pattern classification techniques in the lower-dimensional eigenspace (classification phase).

1.2 Purpose and Contributions of This Paper

This paper is a significant extension of an earlier and much shorter version presented in [16]. The main purpose and contributions of the current paper are summarized as follows:

- We attempt to develop a simple and effective method for gait-based human identification using silhouette analysis. Here, we apply the traditional PCA method like [11], [31] to reduce the dimensionality of the input feature space.
- The proposed method not only analyzes spatiotemporal motion pattern of gait, but also derives a compact statistical description of gait as a continuum. So, this implicitly captures both structural (appearances) and transitional (dynamics) characteristics of gait.
- An integrated background subtraction procedure is proposed. It combines some effective techniques demonstrated in different change detection approaches and obtains smoother gait detection results on the whole (which are very critical to gait analysis).
- Instead of silhouette images usually used in existing silhouette-based work, here we analyze silhouette boundary (i.e., outer contour) and further convert it into an associated 1D signal. This greatly reduces the computational cost of the subsequent processes.
- A new gait period analysis method is proposed. It is believed that this method is suitable to analyze gait frequencies of arbitrary image sequences including a linear walking pattern taken from different views (which is very helpful to time aligning among different gait sequences).
- Compared with most previous algorithms and earlier databases with only a lateral view, performance evaluation in this paper is performed on a database with three views and with size similar to those of most gait databases currently in use. The proposed algorithm is also evaluated on a likely largest data set described by Phillips et al. [33], [36].
- A large number of papers in the literature reported good recognition results on a limited-size database, but none of them made informed comparisons among different algorithms. Here, we provide some quantitative comparative experiments to examine the performance of the proposed algorithm and other recent approaches.
- The proposed method has several desirable properties:

  1. it is easy to comprehend and implement,
  2. it is insensitive to the color and texture of cloth as a silhouette-based approach,
  3. some additional features related to pace, stride, and build are used to improve recognition accuracy, and
  4. experimental results demonstrate that it has a relatively low computational cost.

The remainder of this paper is organized as follows: Section 2 describes related work in order to put ours in context. Section 3 introduces feature extraction, involving human detection and tracking, silhouette shape representation, the training and projection process based on PCA, etc. Gait recognition based on the standard pattern classification techniques is discussed in Section 4. Experimental results are presented and analyzed in Section 5. Section 6 concludes the paper.

2 Related Work

Because human ambulation is one form of human movements, gait recognition is closely related to vision methods that detect, track, and analyze human behaviors in human motion analysis [35]. There has been a rich amount of work on human modeling and tracking [17], [25]. However, the vision researchers have only recently begun to investigate gait as a new biometric feature. Although gait recognition is a newly emergent research area, there have been some attempts in the literature [3], [4], [5], [6], [7], [8], [11], [12], [13], [14], [19], [20], [22], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36]. Current gait recognition approaches may be explicitly classified into two main classes, namely model-based methods and motion-based methods.

Model-based approaches [28], [4], [8], [12], [14], [19], [32] aim to explicitly model human body or motion, and they usually perform model matching in each frame of a walking sequence so that the parameters such as trajectories are measured on the model. An early such attempt [4] modeled gait an articulated pendulum and used the dynamic Hough transform to extract the line representing the thigh in each frame. Fourier analysis was performed on the inclination data of the thigh and phase-weighted magnitude spectra formed gait signatures for recognition. Johnson and Bobick [14] used activity-specific static body parameters for gait recognition without directly analyzing gait dynamics. Yam et al. [19] first tried the running action of gait to recognize people as well as walking. Later, they further explored the intimate relationship between walking and running that was expressed as a mapping based on the idea of phase modulation [28]. The effectiveness of model-based methods, especially in body structure/motion modeling and parameter recovery from a walking video, is still limited allowing for current imperfect vision techniques (e.g., tracking and localizing human body accurately in 2D or 3D space has been a long-term challenging and unsolved problem). Further, the computational cost of model-based methods is relatively high.

Most existing motion-based approaches can be further divided into two main classes [11]. The first class, state-space methods [11], considers gait motion to be composed of a sequence of static body poses and recognizes it by considering temporal variations of observations with respect to those static poses [6], [7], [16], [26], [27]. Using human shapes and their temporal change, Murase and Sakai [6] presented a template matching method based upon eigenspace representation to distinguish different gaits. Huang et al. [7] extended the approach of Murase and Sakai [6] by adding canonical analysis. HMMs (*Hidden Markov Models*) have also been successfully used in gait recognition [26], [27]. The second class, spatiotemporal methods [11], generally characterizes the spatiotemporal distribution generated by gait motion in its continuum [3], [4], [5], [11], [30], [33], [35], [36]. The earliest approach to recognizing walking figures was probably due to Niyogi and Adelson [3]. They distinguished different walkers through extracting their spatiotemporal gait patterns obtained from the curve-fitted "snake." Little and Boyd [5] used frequency and phase features from optical flow information of walking figures to recognize individuals. BenAbdelkader et al. [11], [35] used image self-similarity plots of a moving person to recognize gait. More recently, Hayfron-Acquah et al. [30] described an automatic gait recognition method using the spatiotemporal symmetry, Vega and Sarkar [31] analyzed gait by exploiting nonstationarity in the distribution of feature relationships, and Phillips et al. [33], [36] described a much larger and more challenging database, and presented results for a baseline algorithm for the gait identification problem.

A number of approaches have already shown that it is possible to recognize people by gait. Compared with other widely used biometric features such as face and fingerprint, gait recognition is in its infancy. Most existing approaches are based on some simplifying assumptions, e.g., people walk frontal-parallel to a fixed camera with a plain background and with no occlusion. Furthermore, performance evaluation is usually performed on a relatively small database due to the lack of an accredited common gait database of reasonable size. Vision-based gait recognition for human identification will thus offer us an interesting research topic.

3 Feature Extraction

Before training and recognition, each image sequence including a walking figure is converted into an associated temporal sequence of distance signals at the preprocessing stage.

3.1 Human Detection and Tracking

Human detection and tracking is the first step to gait analysis. Although it is not a main part of our work, in contrast to gait signature extraction and recognition, we still give a detailed introduction for completeness. To extract and track moving silhouettes of a walking figure from the background image in each frame, the change detection and tracking algorithm is adopted which is based on background subtraction and silhouette correlation. The main assumption made here is that the camera is static, and the only moving object in video sequences is the walker. Although this integrated method basically performs well on our data set, it should be noted that robust motion detection in unconstrained environments is an unsolved problem for current vision techniques because it concerns a number of difficult issues such as shadows and motion clutter.

3.1.1 Background Modeling

Background subtraction has been widely used in foreground detection where a fixed camera is usually used to observe dynamic scenes. How to reliably generate the background image from video sequences is critical. Here, the LMedS (Least Median of Squares) method [9] is used to construct the background from a small portion of image sequences even including moving objects. Let I represent a sequence including N images. The resulting background b_{xy} can be computed by [9]

$$b_{xy} = \min_{p} med_t (I_{xy}^t - p)^2, \tag{1}$$

where p is the background brightness value to be determined for the pixel location (x, y), *med* represents the median value, and t represents the frame index ranging

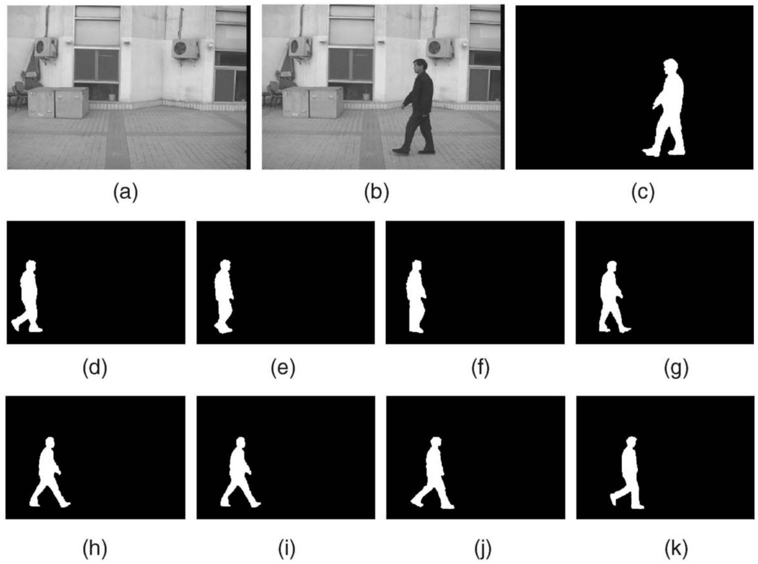


Fig. 2. Examples of moving silhouette extraction and tracking: (a) the background image constructed by the LMedS method, (b) an original image, (c) the extracted silhouette from (b), and (d)-(k) temporal changes of moving silhouettes in a gait pattern (frame 17 to frame 24).

within $1-N$. It is found that N over 60 is sufficient for our data set to generate a reliable background.

3.1.2 Differencing

The brightness change is often obtained through differencing between the background and current image. However, the selection of a suitable threshold for binarization is very difficult, especially in the case of low contrast images as most of moving objects may be missed out since the brightness change is too low to distinguish regions of moving objects from noise [10]. To solve this problem, we use the following extraction function to indirectly perform differencing [10]

$$f(a,b) = 1 - \frac{2\sqrt{(a+1)(b+1)}}{(a+1)+(b+1)} \cdot \frac{2\sqrt{(256-a)(256-b)}}{(256-a)+(256-b)}, \quad (2)$$

where $a(x,y)$ and $b(x,y)$ are the brightness of current image and the background at the pixel position (x,y), respectively, $0 \leq a(x,y), b(x,y) \leq 255$, $0 \leq f(a,b) < 1$. This function can detect the change sensitivity of the difference value according to the brightness level of each pixel in the background image [10].

For each image I_{xy}, the distribution of the above extraction function $f(a(x,y), b(x,y))$ over x and y can be easily obtained. Then, the moving pixels can be extracted by comparing such a distribution against a threshold value decided by the conventional histogram method.

3.1.3 Postprocessing and Tracking

It should be noted that the above process is independently performed for each component R, G, and B in an image. For a given pixel, if one of the three components determines it as the changing point, then it will be set to the foreground. This produces a mask of a region of interest for further processing.

No change detection algorithm is perfect. Hence, it is imperative to remove as much noise and distortion as possible from the segmented foreground. Morphological operators such as erosion and dilation are first used to further filter spurious pixels, and small holes inside the extracted silhouettes are then filled. A binary connected component analysis is finally applied to extract a single highly compact connected region with the largest size.

To eliminate the inaccuracy due to segmentation error, each foreground region is then tracked from frame to frame by a simple correspondence method based on the overlap of their respective bounding boxes in any two consecutive frames [23]. That is, we perform a binary edge correlation between the current and previous silhouette profiles over a small set of displacements [23]. An example of motion segmentation and the tracking process are shown in Fig. 2, from which we can see that the human detection and tracking procedure performs well on our data as a whole. It absolutely does not affect the following feature extraction process though there are a small portion of silhouette distortions such as partial missing of body parts (e.g., invisible arms in Figs. 2d, 2j, and 2k) and the cross of two slightly separated legs (e.g., in Fig. 2f).

3.2 Silhouette Representation

An important cue in determining underlying motion of a walking figure is temporal changes of the walker's silhouette. To make the proposed method insensitive to changes of color and texture of clothes, we use only the binary silhouette. Additionally, for the sake of computational efficiency, we convert these 2D silhouette changes into an associated sequence of 1D signals to approximate temporal pattern of gait. This process is illustrated in Fig. 3.

After the moving silhouette of a walking figure has been tracked, its outer contour can be easily obtained using a border following algorithm. Then, we may compute its shape centroid (x_c, y_c). By choosing the centroid as a reference origin, we unwrap the outer contour counterclockwise to turn it into a distance signal $S = \{d_1, d_2, \ldots, d_i, \ldots, d_{Nb}\}$ that is composed of all distances d_i between each boundary pixel (x_i, y_i) and the centroid

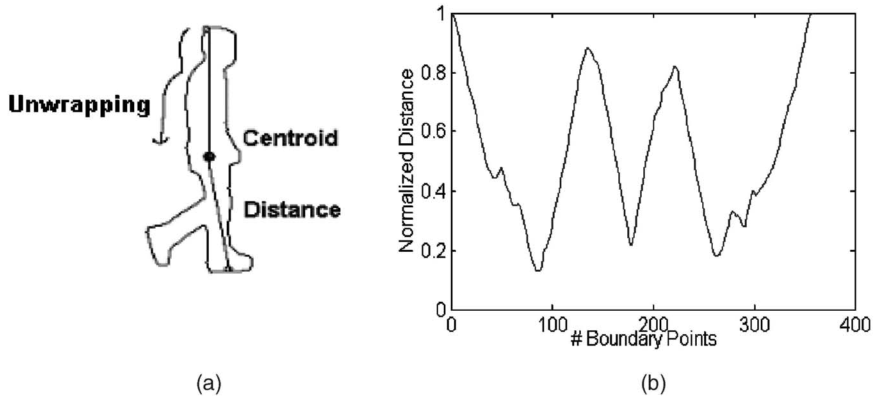


Fig. 3. Silhouette representation: (a) illustration of boundary extraction and counterclockwise unwrapping and (b) the normalized distance signal consisting of all distances between the centroid and the pixels on the boundary.

$$d_i = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2}. \tag{3}$$

This signal indirectly represents the original 2D silhouette shape in the 1D space.

To eliminate the influence of spatial scale and signal length, we normalize these distance signals with respect to magnitude and size. First, we normalize its signal magnitude through $\mathrm{L_1\text{-}norm}$. Then, equally spaced re-sampling is used to normalize its size into a fixed length (360 in our experiments). Additionally, we regularize the walking direction of sequences taken from the same view based upon the symmetry of gait motion during shape representation (e.g., from left to right for all sequences with lateral view). By converting such a sequence of silhouette images into an associated sequence of 1D signal patterns, we will no longer need to cope with those likely noisy silhouette data.

3.3 Training and Projection

3.3.1 PCA Training

The purpose of PCA training is to obtain several principal components to rerepresent the original gait features from a high-dimensional measurement space to a low-dimensional eigenspace. The training process similar to [7] is illustrated as follows:

Given s classes for training, and each class represents a sequence of distance signals of one subject's gait. Multiple sequences of each person can be freely added for training. Let $D_{i,j}$ be the $j\mathrm{th}$ distance signal in class i and N_i the number of such distance signals in the $i\mathrm{th}$ class. The total number of training samples is $N_t = N_1 + N_2 + \ldots + N_s$, and the whole training set can be represented by $[D_{1,1}, D_{1,2}, \ldots, D_{1,N_1}, D_{2,1}, \ldots, D_{s,N_s}]$. We can easily obtain the mean m_d and the global covariance matrix Σ of such a data set by

$$m_d = \frac{1}{N_t} \sum_{i=1}^{s} \sum_{j=1}^{N_i} D_{i,j}, \tag{4}$$

$$\sum = \frac{1}{N_t} \sum_{i=1}^{s} \sum_{j=1}^{N_i} (D_{i,j} - m_d)(D_{i,j} - m_d)^T. \tag{5}$$

If the rank of the matrix Σ is N, then we can compute N nonzero eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_N$ and the associated eigenvectors $e_1, e_2, \ldots, e_N$ based on SVD (*Singular Value Decomposition*).

Generally speaking, the first few eigenvectors correspond to large changes in training patterns. Therefore, for the sake of memory efficiency in practical applications, we may ignore those small eigenvalues and their corresponding eigenvectors using a threshold value T_s

$$W_k = \sum_{i=1}^{k} \lambda_i \Big/ \sum_{i=1}^{N} \lambda_i > T_s, \tag{6}$$

where W_k is the accumulated variance of the first k largest eigenvalues with respect to all eigenvalues. In our experiments, T_s is chosen as 0.95 for obtaining steady results.

3.3.2 Projection

Taking only the $k < N$ largest eigenvalues and their associated eigenvectors, the transform matrix $E = [e_1, e_2, \ldots, e_k]$ can be constructed to project an original distance signal $D_{i,j}$ into a point $P_{i,j}$ in the k-dimensional eigenspace.

$$P_{i,j} = [e_1 \;\; e_2 \;\; \ldots \;\; e_k]^T D_{i,j}. \tag{7}$$

Accordingly, a sequential movement of gait can be mapped into a manifold trajectory in such a parametric eigenspace.

It is well-known that k is usually much smaller than the original data dimension N. That is to say, eigenspace analysis can drastically reduce the dimensionality of input samples. For each training sequence, the projection centroid C_i in the eigenspace is accordingly given by averaging all single projections corresponding to each frame in the sequence.

$$C_i = \frac{1}{N_i} \sum_{j=1}^{N_i} P_{i,j}. \tag{8}$$

4 Recognition

Gait recognition is a traditional pattern classification problem which can be solved by measuring similarities

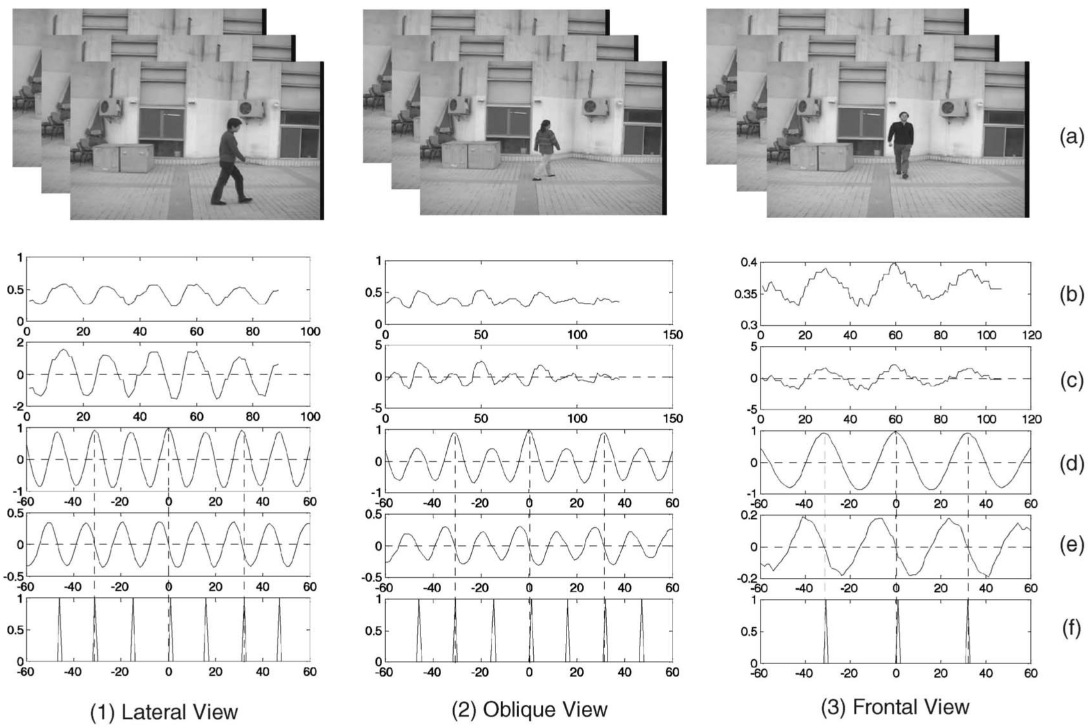


Fig. 4. Gait period analysis: (a) input sequences, (b) aspect ratio signals (i.e., width/height in Fig. 10a) of moving silhouettes, (c) signals after removing the background, (d) autocorrelation signals, (e) first-order derivative signals of autocorrelations, and (f) the positions of peaks.

between reference patterns and test samples in the parametric eigenspace.

4.1 Similarity Measures

4.1.1 Spatiotemporal Correlation

Gait is a kind of spatiotemporal motion pattern, so we use STC (*Spatial-Temporal Correlation*, an extension of 2D image correlation to 3D correlation in the space and time domain [6]) to better capture its spatial structural and temporal transitional characteristics.

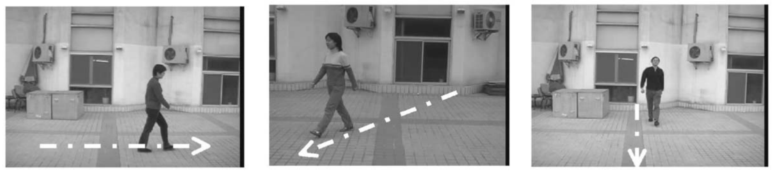
For two input sequences, we can first convert them into a sequence of distance signal $I_1(t)$ and $I_2(t)$ at the preprocessing stage as described in Section 3. Then, they are respectively projected into a trajectory $P_1(t)$ and $P_2(t)$ in the eigenspace using (7). The similarity measure between two such input vector sequences can be computed by [6]

$$d^2 = \min_{ab} \sum_{t=1}^{T} ||P_1(t) - P_2'(at+b)||^2, \tag{9}$$

where $P_2'(at+b)$ is a dynamic time warping vector from $P_2(t)$ with respect to time stretching and shifting for an approximation of the temporal alignment between the two sequences. The selection of the parameters a and b depends on the relative stride frequency and phase difference within a stride (two steps), respectively. Let f_1 and f_2 denote the frequencies of the two gait sequences, then $a = f_2/f_1$. By cropping a subsequence of length f_2 from the second sequence vector repeatedly and stretching it with a, we may obtain its correlation with $P_1(t)$. The average minimum of all prominent valleys of the correlation results determines their similarity.

Gait period analysis has been explored in previous work [20], [22], which serves to determine the frequency and phase of each observed sequence so as to align sequences before matching. Note that a step is the motion between successive heel strikes of opposite feet and that a complete gait period is comprised of two steps. In [20], width time signal of the bounding box of moving silhouette derived from an image sequence is used to analyze gait period. In [22], either width time signal or height time signal is used because the silhouette width for frontal views is less informative, but the silhouette height as a function of time plays an analogous role in periodicity [22]. Different from them, here we choose the aspect ratio of the bounding box of moving silhouette as a function of time so as to enable it to cope effectively with both lateral view and frontal view.

The process of period analysis of each gait sequence similar to [20] is shown in Fig. 4. For an input sequence (Fig. 4a), once the person has been tracked for a certain number of frames, its spatiotemporal gait parameters such as the aspect ratio signal of the moving silhouette can be estimated (Fig. 4b). We may remove its background component by subtracting its mean and dividing by its standard deviation, and then smooth it with a symmetric average filter (Fig. 4c). Further, we compute its autocorrelation to find peaks (Fig. 4d). Finally, we compute its first-order derivative (Fig. 4e) to find peak positions by seeking the positive-to-negative zero-crossing points (Fig. 4f). Due to the bilateral symmetry of human gait, the autocorrelation will sometimes have minor peaks half way between each pair of major peaks [20]. Hence, we estimate the real period as the average distance between each pair of consecutive major peaks. This process has been demonstrated to be computationally feasible with respect to our background subtraction results.



(a) (b) (c)

Fig. 5. Some sample images in the *NLPR* gait database: (a) lateral view, (b) oblique view, and (c) frontal view.

4.1.2 The Normalized Euclidean Distance

Note that the computational cost will increase quickly if the comparison is performed in the spatiotemporal domain, especially when time stretching and shifting is taken into account [6]. Here, we turn to use the NED (*Normalized Euclidean Distance*) between the projection centroids of two gait sequences for the similarity measure to eliminate such matching problems.

Assuming that the trajectories of any two sequences in the eigenspace are $P_1(t)$ and $P_2(t)$, respectively, we can easily obtain their associated projection centroids C_1 and C_2 using (8). Each projection centroid implicitly represents a principal structural shape of certain subject in the eigenspace. The normalized Euclidean distance between the two sequential projection centroids can be defined by

$$d^2 = \left\| \frac{C_1}{\|C_1\|} - \frac{C_2}{\|C_2\|} \right\|^2. \tag{10}$$

Furthermore, for multiple sequences of the same subject, we may also obtain its exemplar projection centroid by further averaging the projection centroids of those single sequences as a reference template for that class. This exemplar centroid will also be used for gait classification in our experiments.

4.2 Classifier

The classification process is carried out via two simple different classification methods, namely the nearest neighbor classifier (*NN*) and the nearest neighbor classifier with respect to class exemplars (*ENN*) derived from the mean projection centroid of those training sequences for a given subject.

Let T represent a test sequence and R_i represent the ith reference sequence. We may classify this test sequence into class c that can minimize the similarity distance between the test sequence and all reference patterns by

$$c = \arg\min_i d_i(T, R_i), \tag{11}$$

where d is the similarity measure described in Section 4.1. Note that d can only choose NED if *ENN* is used. No doubt, a more sophisticated classifier could be employed, but the main interest here is to evaluate the genuine discriminatory ability of the extracted features in our method.

5 Experiments

Extensive experiments are carried out to verify the effectiveness of the proposed algorithm. The following describes the details of the experiments.

5.1 Data Acquisition

A new gait database, called the *NLPR* database, is established for our experiments. A digital camera (Panasonic NV-DX100EN) fixed on a tripod is used to capture gait sequences on two different days in an outdoor environment. All subjects walk along a straight-line path at free cadences in three different views with respect to the image plane, namely, laterally ($0°$), obliquely ($45°$), and frontally ($90°$). The resulting *NLPR* database includes 20 subjects and four sequences for each viewing angle per subject. For instance, when the subject is walking laterally to the camera, the direction of walking is from left to right for two of the four sequences, and from right to left for the remaining. The database therefore includes a total of 240 gait sequences ($20 \times 4 \times 3$). These sequence images with 24-bit full color are captured at a rate of 25 frames per second and the original resolution is 352×240. The length of each image sequence varies with the pace of the walker, but the average is about 90 frames. To the best of our knowledge, our database is probably one of the concurrent gait databases available in the public domain, which is reasonably sized (see Table 3 for a summary of major gait databases currently in use). Some sample images are shown in Fig. 5, where the white line with arrow represents the walking path.

5.2 Preprocessing and Training

For each image sequence, we perform motion segmentation and tracking to extract the walking figure from the background image as described in Section 3.1. Those extracted 2D silhouettes are accordingly converted into an associated sequence of 1D distance signals before training and projection as presented in Section 3.2.

We choose a small portion of such distance signal sequences including all classes for training. The training process based on PCA is accomplished in the manner described in Section 3.3. We keep the first 15 eigenvalues and their associated eigenvectors to form the eigenspace transformation matrix. Fig. 6 gives the first three eigenshapes for each viewing angle. From Fig. 6, we can see that these eigen-curves are either odd symmetric or even symmetric, which reveals that gait has a characteristic of symmetry.

Once the eigenspace is obtained, each distance signal derived from each silhouette image can be represented by a linear combination of these 15 principal eigenvectors. That is, each distance signal can be mapped into one point in a 15-dimensional eigenspace. Each gait sequence will be accordingly projected into a manifold trajectory in the eigenspace. The projection trajectories of three trained

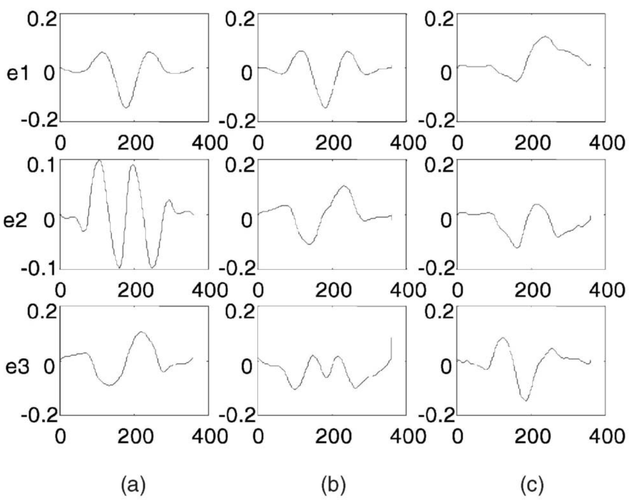


Fig. 6. The first three eigenvectors for each viewing angle obtained by PCA training: (a) lateral view, (b) oblique view, and (c) fontal view.

sequences with respect to lateral view, oblique view, and frontal view, respectively, are shown in Fig. 7, where only three-dimensional eigenspace is used for visualization.

5.3 Results and Analysis

5.3.1 Identification Mode

A useful classification performance measure was introduced by the FERET protocol for the evaluation of face recognition algorithms [24]. It is defined as the cumulative probability $p_{(k)}$ that the real class of a test measurement is among its top k matches [24]. The performance statistics are reported as the cumulative match scores. The rank k is plotted along the horizontal axis, and the vertical axis is the percentage of correct matches [24].

Here, we use the leave-one-out cross-validation rule with the *NLPR* database to estimate the performance of the proposed method. Each time we leave one image sequence out as a test sample and train on the remainder. After computing the similarity differences between the test sample and the training data, the *NN* or *ENN* is then applied for classification. Fig. 8 shows the cumulative match scores for ranks up to 20, where Fig. 8a uses the STC similarity measure and Fig. 8b uses the NED similarity measure with respect to projection centroids (solid line) and exemplar projection centroids (dotted line), respectively. It is noted that the correct classification rate is equivalent to $p_{(1)}$ (i.e., Rank=1). That is, for side view, oblique view, and frontal view, the correct classification rates are, respectively, 65, 63.75, and 77.5 percent with *NN* and STC, 65, 66.25, and 85 percent with *NN* and NED, and 75, 81.25, and 93.75 percent with *ENN* and NED.

From Fig. 8, we can draw the following conclusions:

- The identification performance using NED is, in general, better than that using STC. In theory, STC can better capture spatiotemporal characteristics of gait motion than NED, and it is expected to obtain better recognition accuracy. Such experimental results are probably due to the fact that the segmentation errors in each frame caused by either noise or clothing dithering when walking in different gait sequences may be accumulated into a quick-enlarged match error to some extent during spatiotemporal correlation using direct frame-frame match. However, the average projection over a whole temporal sequence can provide a powerful method for overcoming noise in individual frames. We hypothesize that a more robust method for silhouette extraction would yield an improvement in STC scores.
- The NED based on the exemplar projection centroid performs better than NED using only a single projection centroid. For each subject, although his or her gaits at different times are almost perceived as the same, there are slightly changes between them. The average of multiple sample sequences may serve to provide a more standard gait pattern for that specific person than only a single and random sample sequence.
- The recognition performance under frontal walking (90°) is the best. This is probably not what happens with most algorithms. The proposed method in essence implicitly captures more structural appearance information of body biometrics. Therefore, this result is probably due to the averaging associated

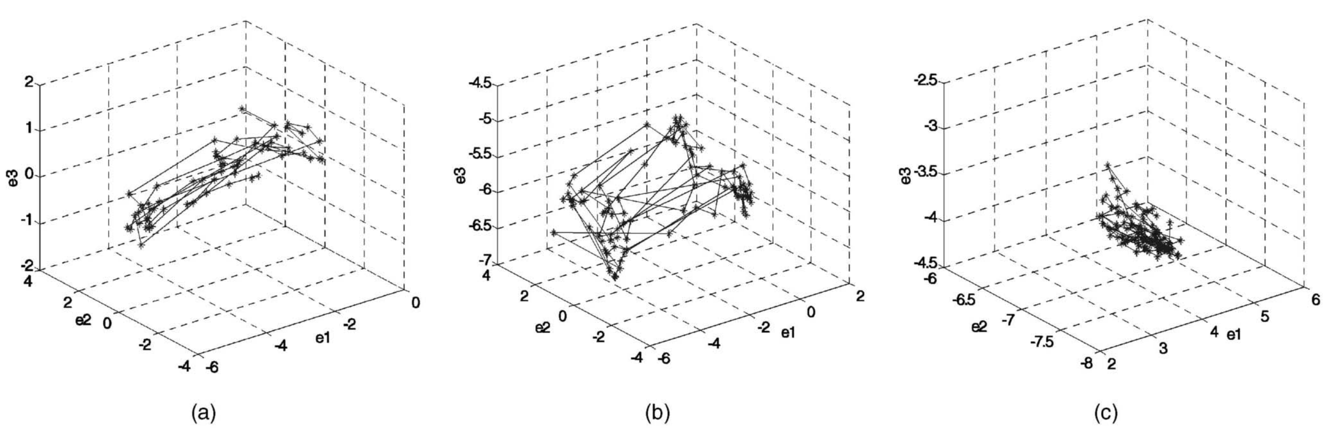
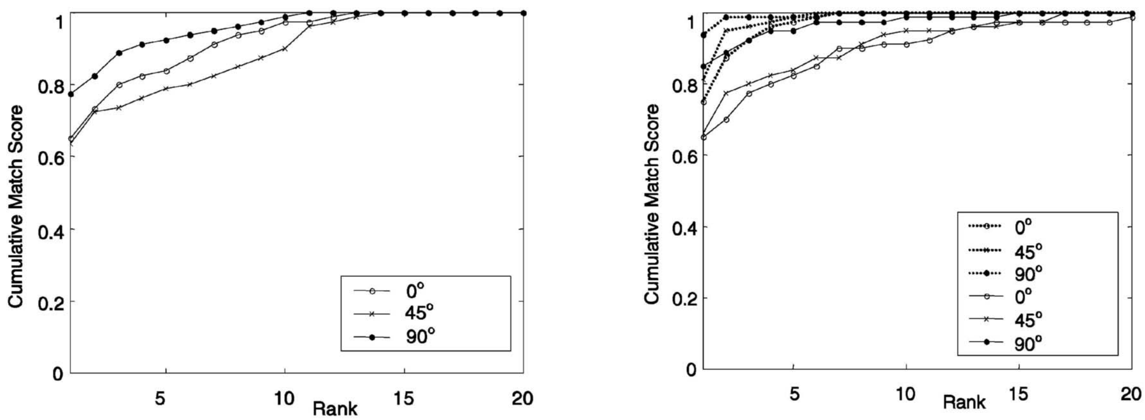


Fig. 7. The projection trajectories of three training gait sequences (only the three-dimensional eigenspace is used here for clarity): (a) lateral view, (b) oblique view, and (c) frontal view.



(a) (b)

Fig. 8. Identification performance based on the cumulative match scores: (a) classifier based on STC and (b) classifiers based on NED with respect to single projection centroid (solid line) and exemplar projection centroid (dotted line), respectively.

with the silhouette shape analysis because there are less severe variations of silhouette appearances in such gait patterns compared with the other views.

5.3.2 Verification Mode

For completeness, we also estimate FAR (*False Acceptance Rate*) and FRR (*False Reject Rate*) via the leave-one-out rule in verification mode. That is, we leave one example out, train the classifier using the remaining, and then verify the left-out sample on all 20 classes. Note that, in each of these 80 iterations for each viewing angle, there is one genuine attempt and 19 imposters since the left-out sample is known to belong to one of the 20 classes. Fig. 9 shows the ROC (*Receiver Operating Characteristic*) curve using the NED similarity measure with exemplar projection centroids, from which we can see that the EERs (*Equal Error Rate*) are about 20, 13, and 9 percent for 0, 45, and 90 degree views, respectively. Here, the verification performance of frontal view is also better than those of the other views.

5.3.3 Validation Based on Physical Features

In experiments, we find that recognition errors often occurred when the two smallest values of the similarity function are very close. That is, the classifier recognizes a test sample as the class with the smallest similarity value, while the true classification should be the class with the second smallest similarity value. Therefore, when the absolute difference between the last two minima is lower than a predefined threshold, we use some additional features available from the training sequences to validate the final decision. These features may be pace, body height, build, and stride length, some of which have been independently used for personal identification in previous algorithms [8], [14], [20]. Because we do not perform camera calibration when establishing the *NLPR* database, we cannot extract real parameters in the world coordinate like [20]. To avoid the effect of foreshortening in terms of different views, here we only take image sequences from the lateral view as test examples to examine the effectiveness of such validation. The walking figures from the lateral-view images have approximately the same depth, and it is believed that static body parameters recovered from such a single view produce high discrimination power [8].

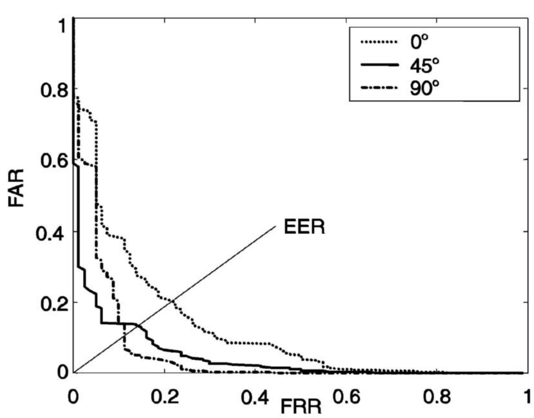


Fig. 9. ROC curves of gait classifier based on NED with respect to three viewing angles.

Extraction of additional features is illustrated in Fig. 10. For each frame, a bounding box is placed around the tracked silhouette, and we can easily obtain some parameters such as its centroid, width and height. To extract the other body points, the vertical positions of the chest and ankles for a body height H are estimated by an anatomical study [21] to be $0.720H$ and $0.039H$, respectively. Further, we can obtain its chest width and the distance between the left foot and the right foot by calculating the horizontal coordinate from two border points (note that, for ankles, we choose the leftmost and the rightmost points along the horizontal axis). Once the person has been tracked for a certain number of frames, the spatiotemporal trajectories of these gait parameters can be derived. We select the following additional features to describe aspects of pace, stride, and build. The first parameter, gait period T, can be obtained as mentioned in Section 4.1. From the *NLPR* database, we find that the gait period usually ranges from 26 to 37 frames. The second parameter is stride SL which is measured only at the maximal separation point of the feet during the double support phase of a gait cycle. The last two parameters,

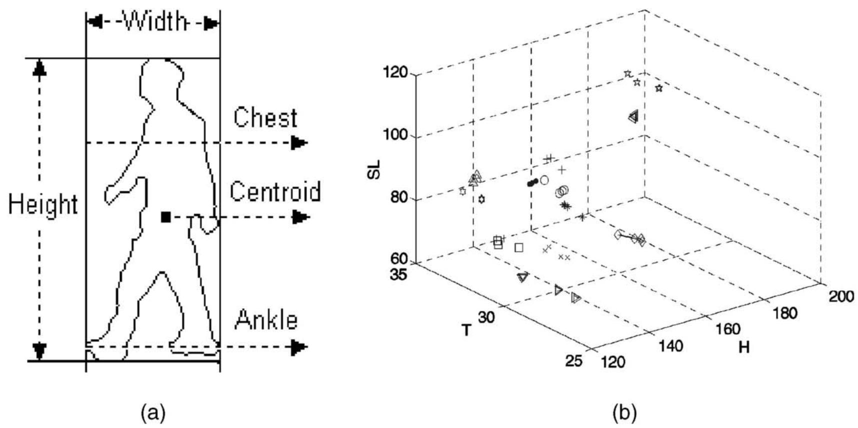


Fig. 10. Extraction of additional features: (a) illustration of human silhouette and (b) a distribution of physical features in 3D space (note that SL and H are measured by pixel distance).

TABLE 1
The Effectiveness of Validation Based on Physical Features (the *NLPR* Database ($0°$))

Similarity Measure	Classifier	CCR (No validation)	CCR (With validation)
STC	*NN*	65.00%	68.75%
NED	*NN*	65.00%	70.00%
NED	*ENN*	75.00%	82.50%

namely the body height H and the ratio R between the chest width and the body height, are used to reflect build information (tall versus short and thin versus fat). We obtain these parameters by averaging measurements over each time instance, namely the time instance with the minimal separation points between two feet (when the subject is at his maximal height) and the time instance with the maximal separation points between two feet (when the subject have the maximal stride). These parameters are finally concatenated to form a four-dimensional vector $< T, SL, H, R >$ for each sequence. Fig. 10b shows a distribution of such physical features in 3D space for clarity, where the same markers represent the results of different sequences of the same subject. From Fig. 10b, we can see that an efficient combination of these physical properties will bring considerable discriminatory power to gait classification.

The effect of the above validation procedure on the CCR (*Correct Classification Rate*) is shown in Table 1, from which we can see that the recognition performance after the inclusion of the above additional features is indeed improved. If we can construct a depth conversion factor as a function of the depth of the subject from the camera as described in [14], or use camera calibration to convert distances measured in the image from pixels to world units, we may extend the validation step to solve the effects of foreshortening in terms of other views. Although the results are very encouraging, further experiments on a larger database need to be investigated in order to be more conclusive.

5.4 Comparisons

The lack of a common database (e.g., the FERET database in face recognition) and evaluation methodology has been an apparent limitation in the development of gait recognition algorithms. It is exciting to see that Phillips et al. [33] are trying to address this problem by setting up a standard data set to measure or determine what factors affect performance. It is expected that they extend their work on establishing a more challenging database and proposing a more standard baseline method for development and evaluation of new algorithms. As we know, a large number of papers in the literature reported good recognition results on a limited-size database, none of which made informed comparisons among different algorithms. To examine the performance of the proposed algorithm, we provide two comparative experiments as follows.

5.4.1 Comparison 1

Here, we first compare the performance of the proposed algorithm with that of a closely related method described in [11]. This approach named Eigengait is based on PCA, and the major difference from our method is that it uses image self-similarity plots as the original measurements. This algorithm was evaluated on a data set of Little and Boyd [5] and achieved a recognition rate of 80, 82.5, and 90 percent with respect to $k = 1$, 3, and 5 using the k-nearest neighbor classifier. We reimplement this method using the *NLPR* database with a lateral viewing angle. The best recognition rate is 72.5 percent (see Table 2), which is a little lower than our method even with no validation (75.00 percent). Due to the lack of the database used in [5], here we are unable to test the proposed algorithm on the data set.

We also compare the performance of the proposed algorithm with those of a few recent silhouette-based methods described in [22], [34], and [36], respectively. To some extent, they reflect the latest and best work of these research groups in gait recognition. Based on body shape

TABLE 2
Comparison of Several Recent Algorithms on the *NLPR* Database (0°)

Methods	Top 1 (%)	Top 5 (%)	Top 10 (%)	Computational cost (min/seq)
BenAbdelkader 2001 [11]	72.50	88.75	96.25	Medium (8.446)
Collins 2002 [22]	71.25	78.75	87.50	High (17.807)
Lee 2002 [34]	87.50	98.75	100	Low (2.365)
Phillips 2002 [36]	78.75	91.25	98.75	Highest (200)
Our method (No/with validation)	75.00/82.50	97.50/100	100/100	Lowest (2.054)

and gait, Collins et al. [22] established a method based on template matching of body silhouettes in key frames for human identification. Lee et al. [34] described a moment-based representation of gait appearance for the purpose of person identification. Phillips et al. [36] proposed a baseline algorithm for human identification using spatiotemporal correlation of silhouette images. Here, we reimplement these methods using the same silhouette data from the *NLPR* database with lateral view. Based on the FERET protocol with rank of 1, 5, and 10, the best results of all algorithms are summarized in Table 2, from which we can see that our method compares favorably with others, and outperforms Phillips et al. [36] and Collins et al. [22]. We also found that the computational cost of [22] and [36] was much higher than that of [11], [34] and our method. Here, the listed computational cost is an approximately average consumed time for each test sequence using Matlab 6.1 on a PIII processor working on 733Mhz with 256Mb DRAM (note that this process only includes feature extraction and matching, and excludes gait segmentation and the training phase).

The results here only provide preliminary comparative performance and may not be generalized to say that a certain algorithm is always better than others. Each method might have its own unique advantages and disadvantages under different testing conditions, so further evaluations and comparisons on more realistic and challenging databases are needed.

5.4.2 Comparison 2

One might argue that comparisons with recognition rates on different databases are not very valuable, and think that the assessment of factors in real gait recognition is reasonably demanding. That is, the real question is how the results will generalize to larger data sets under more real-world conditions [33], [36]. So, it will be more meaningful to test the proposed method to show robustness with respect to different factors potentially affecting performance.

An overview of some typical databases used in the literature is listed in Table 3. In the following, we are going to investigate the effectiveness of the extracted gait features with respect to different variations on the "gait challenge" data set described by Phillips et al. [33], [36], as this database is the largest available to date in terms of number of people, number of video sequences, and the variety of conditions under which a person's gait is collected. Some samples are given in Fig. 11. So far, this data set available for gait analysis consists of 452 sequences from 74 subjects walking in elliptical paths on two surface types (**C**oncrete and **G**rass), from two camera viewpoints (**L**eft and **R**ight), and with two shoe types (**A** and **B**). Thus, we have eight possible different conditions for each person (Refer to Table 4) [33], [36].

For fairness in comparison, the following points should be made clear:

- Considering the constraints of processing time and storage space (about 300 GB), we directly used the same silhouette data from the University of South Florida. These data are noisy, e.g., missing of body parts, small holes inside the objects, and severe shadow around feet. Therefore, we make preprocessing to extract a single-connected region with the

TABLE 3
Overview of Some Typical Databases Used in the Literature

Database	UCSD	NLPR	US (1)	CMU	MIT	UMD (1)	UMD (2)	GVU	USF
Environment	O	O	I	I	I	O	O	I or O	O
Walk surface	G1	G1	F	T	F	G1	G1	G1 or F	G2 or C
# Subjects	6	20	12	25	24	25	55	20	74
# Sequences	40	240	48	-	194	100	-	-	452
# Views	1	3	1	6	1	2	2	1	2
Synchronized	N/A	N/A	N/A	Y	N/A	N/A	N	N/A	N
# Walk styles	1	1	1	4	1	1	1	1	1
Frame rate	30	25	25	30	15	20	20	25	30

Note: O-Outdoor, I-Indoor, G1-Ground, G2-Grass, F-Floor, T-Treadmill, C-Concrete

Fig. 11. Some sample images in the database described by Phillips et al. [33], [36].

largest size in each frame to fit the data to our method regardless of severe segmentation errors.

- We put forward a set of challenge experiments in terms of the same gallery and probe sets as the literature [33]. For each experiment, we also measure performance for identification scenario following the pattern of the FERET standard evaluations [24].
- We use the STC similarity measure similar in essence to the baseline algorithm described in [33], [36], with only one difference of interframe similarity representations. Here, we set $a = 1$ without considering different paces among subjects because we cannot effectively perform gait period analysis for a nonlinear (elliptic) walking pattern.

Table 4 lists basic performance indicators of the proposed algorithm in the seven challenge experiments, namely, the identification rate (P_I) for ranks of 1 and 5, where the number of subjects in each subset is in square bracket, and the optimized performance stated by [33] (an extended version of [36]) are in parentheses for comparison. From Table 4, we can draw the same conclusions as [33]. That is, among the three variations, the viewpoint seems to have the least impact and the surface type has the most impact [33]. As a whole, our method is only slightly inferior to [33] in identification rate with the USF database, but far superior in computational cost (see Table 2). As for the identification rate, we think that noisy segmentation results bring about great impact on feature training and extraction to our method, and that an elliptic walking path bring about the challenge for our view-based algorithm (i.e., we set a nearly linear walking path in data acquisition, which is consistent with most past databases and probably more realistic in real cases). The performance of most existing algorithms performing on sequences with a linear walking and side view (e.g., [11], [22], and [34] in Table 3) does naturally result from the serious impact of the above two aspects in both implementation (even may not work effectively) and the resulting performance. As for the computational cost, the baseline algorithm proposed by Phillips et al. [33], [36] essentially belongs to an unlimited temporal-spatial correlation process. Unlike other previous work, it performs intersequence correlation repeatedly using the segmented silhouette images to measure similarity between any two sequences because it does not have the explicit training procedure for extracting a genuine compact feature vector for each sequence.

In summary, Comparison 1 shows that the proposed method outperforms [11], [22], [36] and is comparable to [34] in recognition rate, but is the lowest in computational cost. Comparison 2 demonstrates that our method is comparable to [33] to some extent. As a whole, our method is comparable to the existing approaches in recognition accuracy, but far superior in observed execution times. As stated above, these are just basic comparative results. More detailed statistical studies on larger and more realistic data sets are desirable to further quantify the algorithm performance.

5.5 Discussions and Future Work

To provide a general approach to automatic person identification in unconstrained environments, much remains to be done.

Although our results are valid, we cannot conclude much about gaits. Further evaluation on a much larger and most-varied database is still needed. We are planning to set up such a database with more subjects, more sequences, and more variations in conditions.

The across-day condition represents the hardest test of performance evaluation of a gait recognition method. This is due to same-subject differences caused by changing clothing (bulky versus thin), especially in different seasons. Our method is based on silhouette analysis, so it is

TABLE 4
Basic Results on Phillips et al.'s Database Using the Gallery Set (G, A, R)

Exp.	Probe	Difference	P_I (%), Rank=1	P_I (%), Rank=5
A	(G, A, L) [71]	View	70.42 (79)	92.95 (96)
B	(G, B, R) [41]	Shoe	58.54 (66)	82.92 (81)
C	(G, B, L) [41]	Shoe, View	51.22 (56)	70.73 (76)
D	(C, A, R) [70]	Surface	34.33 (29)	64.18 (61)
E	(C, B, R) [44]	Surface, Shoe	21.43 (24)	45.24 (55)
F	(C, A, L) [70]	Surface, View	27.27 (30)	38.57 (46)
G	(C, B, L) [44]	Surface, Shoe, View	14.29 (10)	26.19 (33)

inevitably subject to the effects of different types of clothes. In fact, most past work is subject to the effect of shape information due to the direct use of motion segmentation results. Creating more reference sequences with different clothes may be of use to solve this problem.

The lack of generality of viewing angles is a limitation to most gait recognition algorithms. Our present method is view dependent like most previous work, so a useful experiment would be to determine the sensitivity of the features to different views, whose results would enable a multicamera tracking system to select an optimal view for recognition [5]. Another obvious way to generalize the algorithm itself is to store training sequences taken from multiple viewpoints and to classify both the subject and the viewpoint [22].

It is more sufficient for recognition to extract dynamic information such as the oscillatory trajectories of joints. Therefore, 3D human body modeling and tracking might prove to be of benefit [32]. Future work may try to combine both static and dynamic features of gait such as posture, arm/leg/hip swing, etc.

Also, seeking better similarity measures, designing more sophisticated classifiers, gait segmentation, and the evaluation of different scenarios deserve more attention in future work.

6 Conclusions

With the increasing demands of visual surveillance systems, human identification at a distance has recently gained more interest. Gait is a potential behavioral feature and many allied studies have demonstrated that it has a rich potential as a biometric for recognition. The development of computer vision techniques has also assured that vision-based automatic gait analysis can be gradually achieved.

This paper has described a simple but effective method for automatic person recognition from body silhouette and gait. The combination of a background subtraction procedure and a simple correspondence method is used to segment and track spatial silhouettes of a walking figure. Simple feature selection and parametric eigenspace representation reduce the computational cost significantly during training and recognition. A large number of experimental results have demonstrated the validity of the proposed algorithm. Although accomplished under some simplified assumptions like previous work, this work has been proven to be an encouraging progress to gait-based human identification.

Acknowledgments

The authors would like to thank Dr. S. Sarkar from the University of South Florida for providing them with the database. Their thanks also go to those who have contributed to the establishment of the *NLPR* database, and to the anonymous referees for their thorough review and constructive comments. This work is supported in part by the National Science Foundation of China (Grant No. 69825105 and 60105002), the Natural Science Foundation of Beijing (Grant No. 4031004), and the National 863 High-Tech R&D Program of China (Grant No. 2002AA117010).

References

[1] http://www.darpa.mil/iao/HID.htm or http://www.humanid.org/index.html, 2000.

[2] S. Stevenage, M. Nixon, and K. Vince, "Visual Analysis of Gait as a Cue to Identity," *Applied Cognitive Psychology,* vol. 13, pp. 513-526, 1999.

[3] S. Niyogi and E. Adelson, "Analyzing and Recognizing Walking Figures in XYT," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* pp. 469-474, 1994.

[4] D. Cunado, M. Nixon, and J. Carter, "Using Gait as a Biometric, via Phase-Weighted Magnitude Spectra," *Proc. Int'l Conf. Audio and Video-Based Biometric Person Authentication,* pp. 95-102, 1997.

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

[6] H. Murase and R. Sakai, "Moving Object Recognition in Eigenspace Representation: Gait Analysis and Lip Reading," *Pattern Recognition Letters,* vol. 17, pp. 155-162, 1996.

[7] P. Huang, C. Harris, and M. Nixon, "Human Gait Recognition in Canonical Space Using Temporal Templates," *IEE Proc. Vision Image and Signal Processing Conf.,* vol. 146, no. 2, pp. 93-100, 1999.

[8] A. Johnson and A. Bobick, "A Multiview Method for Gait Recognition Using Static Body Parameters," *Proc. Int'l Conf. Audio- and Video-Based Biometric Person Authentication,* pp. 301-311, 2001.

[9] Y. Yang and M. Levine, "The Background Primal Sketch: An Approach for Tracking Moving Objects," *Machine Vision and Applications,* vol. 5, pp. 17-34, 1992.

[10] Y. Kuno, T. Watanabe, Y. Shimosakoda, and S. Nakagawa, "Automated Detection of Human for Visual Surveillance System," *Proc. Int'l Conf. Pattern Recognition,* pp. 865-869, 1996.

[11] C. BenAbdelkader, R. Culter, H. Nanda, and L. Davis, "EigenGait: Motion-Based Recognition of People Using Image Self-Similarity," *Proc. Int'l Conf. Audio- and Video-Based Biometric Person Authentication,* pp. 284-294, 2001.

[12] R. Tanawongsuwan and A. Bobick, "Gait Recognition from Time-Normalized Joint-Angle Trajectories in the Walking Plane," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2001.

[13] G. Shakhnarovich, L. Lee, and T. Darrell, "Integrated Face and Gait Recognition from Multiple Views," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2001.

[14] A. Bobick and A. Johnson, "Gait Recognition Using Static, Activity-Specific Parameters," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2001.

[15] M. Nixon, J. Carter, D. Cunado, P. Huang, and S. Stevenage, "Automatic Gait Recognition," *BIOMETRICS Personal Identification in Networked Soc.,* A. Jain, ed., 1999.

[16] L. Wang, W. Hu, and T. Tan, "A New Attempt to Gait-Based Human Identification," *Proc. Int'l Conf. Pattern Recognition,* 2002.

[17] D. Gavrila, "The Visual Analysis of Human Movement: A Survey," *Computer Vision and Image Understanding,* vol. 73, no. 1, pp. 82-98, 1999.

[18] A. Jain, R. Bolle, and S. Pankanti, *Biometrics: Personal Identification in Networked Society.* Kluwer Academic Publishers, 1999.

[19] C. Yam, M. Nixon, and J. Carter, "Gait Recognition by Walking and Running: A Model-Based Approach," *Proc. Asia Conf. Computer Vision,* pp. 1-6, 2002.

[20] C. BenAbdelkader, R. Culter, and L. Davis, "Stride and Cadence as a Biometric in Automatic Person Identification and Verification," *Proc. Int'l Conf. Automatic Face and Gesture Recognition,* 2002.

[21] D. Winter, *The Biomechanical and Motor Control of Human Movement,* second ed. John Wiley & Sons, 1990.

[22] 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,* 2002.

[23] I. Haritaoglu, D. Harwood, and L. Davis, "W^4: Real-Time Surveillance of People and Their Activities," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, no. 8, pp. 809-830, Aug. 2000.

[24] J. Phillips, H. Moon, S. Rizvi, and P. Rause, "The FERET Evaluation Methodology for Face Recognition Algorithms," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, no. 10, pp. 1090-1104, Oct. 2000.

[25] L. Wang, W.M. Hu, and T.N. Tan, "Recent Developments in Human Motion Analysis," *Pattern Recognition,* vol. 36, no. 3, pp. 585-601, 2003.

[26] A. Kale, A. Rajagopalan, N. Cuntoor, and V. Krüger, "Gait-Based Recognition of Humans Using Continuous HMMs," *Proc. Int'l Conf. Automatic Face and Gesture Recognition,* 2002.

[27] Q. He and C. Debrunner, "Individual Recognition from Periodic Activity Using Hidden Markov Models," *Proc. IEEE Workshop Human Motion,* 2000.
[28] C. Yam, M. Nixon, and J. Carter, "On the Relationship of Human Walking and Running: Automatic Person Identification by Gait," *Proc. Int'l Conf. Pattern Recognition,* 2002.
[29] P. Cattin, D. Zlatnik, and R. Borer, "Sensor Fusion for a Biometric System Using Gait," *Proc. Multisensor Fusion and Integration for Intelligent Systems Conf.,* 2001.
[30] J. Hayfron-Acquah, M. Nixon, and J. Carter, "Human Identification by Spatio-Temporal Symmetry," *Proc. Int'l Conf. Pattern Recognition,* 2002.
[31] I. Vega and S. Sarkar, "Experiments on Gait Analysis by Exploiting Nonstationarity in the Distribution of Feature Relationships," *Proc. Int'l Conf. Pattern Recognition,* 2002.
[32] B. Bhanu and J. Han, "Individual Recognition by Kinematic-Based Gait Analysis," *Proc. Int'l Conf. Pattern Recognition,* 2002.
[33] P. Phillips, S. Sarkar, I. Robledo, P. Grother, and K. Bowyer, "The Gait Identification Challenge Problem: Data Sets and Baseline Algorithm," *Proc. Int'l Conf. Pattern Recognition,* 2002.
[34] L. Lee and W. Grimson, "Gait Analysis for Recognition and Classification," *Proc. Int'l Conf. Automatic Face and Gesture Recognition,* pp. 155-162, 2002.
[35] C. BenAbdelkader, R. Cutler, and L. Davis, "Motion-Based Recognition of People in EigenGait Space," *Proc. Int'l Conf. Automatic Face and Gesture Recognition,* pp. 267-274, 2002.
[36] P. Phillips, S. Sarkar, I. Robledo, P. Grother, and K. Bowyer, "Baseline Results for Challenge Problem of Human ID Using Gait Analysis," *Proc. Int'l Conf. Automatic Face and Gesture Recognition,* pp. 137-142, 2002.
[37] G. Bingham, R. Schmidt, and L. Rosenblum, "Dynamics and the Orientation of Kinematic Forms in Visual Event Recognition," *J. Experimental Psychology: Human Perception and Performance,* vol. 21, no. 6, pp. 1473-1493, 1995.
[38] M. Köhle, D. Merkl, and J. Kastner, "Clinical Gait Analysis by Neural Networks: Issues and Experiences," *Proc. IEEE Symp. Computer-Based Medical Systems,* pp. 138-143, 1997.

Liang Wang received the BSc degree in electrical engineering and MSc degree in video processing and multimedia communication from Anhui University, Hefei, China, in 1997 and 2000, respectively. He is currently a PhD candidate in pattern recognition and intelligent systems in the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China. He has published more than 10 papers in major international journals and conferences. His current research interests include computer vision, pattern recognition, digital image processing and analysis, multimedia, visual surveillance, etc.

Tieniu Tan received the BSc degree in electronic engineering from Xi'an Jiaotong University, China, in 1984, and the MSc, DIC, and PhD degrees in electronic engineering from the Imperial College of Science, Technology and Medicine, London, UK, in 1986, 1986, and 1989, respectively. He joined the Computational Vision Group in the Department of Computer Science at The University of Reading, England, in October 1989, where he worked as a research fellow, senior research fellow, and lecturer. In January 1998, he returned to China to join the National Laboratory of Pattern Recognition at the Institute of Automation of the Chinese Academy of Sciences, Beijing, China. He is currently a professor and the director of the National Laboratory of Pattern Recognition as well as president of the Institute of Automation. He has published widely on image processing, computer vision, and pattern recognition. His current research interests include speech and image processing, machine and computer vision, pattern recognition, multimedia, and robotics. He serves as a referee for many major national and international journals and conferences. He is an associate editor of *Pattern Recognition* and *IEEE Transactions on Pattern Analysis and Machine Intelligence*, and the Asia editor of *Image and Vision Computing*. Dr. Tan was an elected member of the Executive Committee of the British Machine Vision Association and Society for Pattern Recognition (1996-1997) and is a founding co-chair of the IEEE International Workshop on Visual Surveillance. He is a senior member of the IEEE.

Huazhong Ning received the BSc degree (2000) in computer science from the Department of Special Class for Gifted Young of University of Science and Technology of China. He is currently a master candidate in the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China. He has published more than six papers on major national journals and international conferences. His main research interests include computer vision, video computing, tracking, human computer interaction, image processing, pattern recognition, graphics, etc.

Weiming Hu received the PhD degree from the Department of Computer Science and Engineering at Zhejiang University, China. He worked as a postdoctoral research fellow at the Institute of Computer Science and Technology, Founder Research and Design Center, Peking University, from April 1998 to March 2000. In April 2000, he joined the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China, as an associate professor. He has published more than 40 papers on major national journals and international conferences. His current research interests include visual surveillance and monitoring of dynamic scenes, recognition and filtering of Internet objectionable images, neural network, computer vision, etc.