Available online at www.sciencedirect.com
SCIENCE DIRECT
Pattern Recognition 39 (2006) 969–979
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/patcog

# Gait recognition using linear time normalization☆

Nikolaos V. Boulgouris[a,*], Konstantinos N. Plataniotis[b], Dimitrios Hatzinakos[b]

[a]*Department of Electronic Engineering, Division of Engineering, King's College London, Strand, London WC2R 2LS, UK*
[b]*The Edward Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Canada*

Received 20 July 2005

**Abstract**

We present a novel system for gait recognition. Identity recognition and verification are based on the matching of linearly time-normalized gait walking cycles. A novel feature extraction process is also proposed for the transformation of human silhouettes into low-dimensional feature vectors consisting of average pixel distances from the center of the silhouette. By using the best-performing of the proposed methodologies, improvements of 8–20% in recognition and verification performance are seen in comparison to other known methodologies on the "Gait Challenge" database.


*Keywords:* Gait; Angular analysis; Time normalization; Recognition; Verification

## 1. Introduction

The analysis of the human way of walking, termed gait, can be used in several applications ranging from medical applications to security. Such applications include the detection of postural disturbances due to mobility disorders or ageing, the rehabilitation of locomotion in stroke and spinal cord injured persons, and the development of gait models for walking machines. The specific use of gait as a biometric trait in recognition/verification applications has recently emerged as a very attractive research area due to its applications on surveillance and identification systems. In this context, gait seems to be an ideal choice since it has the unique characteristic of being unobtrusive, i.e., it can be observed at-a-distance and without requiring the cooperation or consent of the observed subject.

The deployment of gait in identity recognition or verification systems involves a variety of image processing and pattern recognition tasks. Specifically, a gait-based recognition/verification system involves vision-based input capturing, gait feature extraction, and gait recognition/verification. Gait feature extraction is possible using sequences in which the walking person (foreground) has been separated from his/her surroundings (background). This process is often performed imperfectly and this is why an additional denoising process, tailored to the specific task of extracting the walking subject, is needed prior to feature extraction.

Existing approaches for gait feature extraction attempt to analyze gait sequences and capture discriminative information that is subsequently used for recognition/verification purposes. Of particular interest are techniques which try to tackle the gait recognition problem using only sequences of silhouettes. Such techniques do not require texture information which may not be available in practice. In this paper, we propose a methodology for gait analysis using silhouette sequences and then focus our attention on the tasks of recognition and verification using features derived from the preceded analysis technique [1,2].

### 1.1. Gait terminology

Walking is a periodic process with its period termed *gait period* whereas the sequence of stances during a gait period

☆ Parts of this work were presented in Refs. [1,2]. Part of this work was conducted when the first author was with the Electrical and Computer Engineering Department, University of Toronto, Canada.

* Corresponding author. Tel.: +44 20 7848 2361; fax: +44 20 7848 2932.
*E-mail address:* nikolaos.boulgouris@kcl.ac.uk (N.V. Boulgouris).

0031-3203/$30.00 
doi:10.1016/j.patcog.2005.10.013

Fig. 1. Several stances during a gait cycle.

is termed *gait cycle*. A cycle of a silhouette sequence is depicted in Fig. 1. Since in each gait cycle there are two self-similar sequences (one starting with the left leg moving forward and one with the right), the term *half-cycles* is also used in order to describe the two sub-cycles in a gait cycle.

### 1.2. Previous work

In this section we review several approaches for gait recognition. Although direct comparison and evaluation among them is very difficult, in the following we try to highlight their basic features.

Gait recognition was first investigated in Ref. [3] where leg angle signals were extracted from the video sequence in order to be used for identification.

A baseline experiment was carried out at the University of South Florida which was intended to serve as a reference point for gait recognition methods [4]. Specifically, several sets of gait sequences were formed; one of them was considered to be the system database set, whereas the rest of the sets were used as test sets. In the sequel, the similarity among the sequences in the database set and the sequences in each of the test sets were calculated by direct comparison of the frames in the two sequences. The subjects in the database set were then ranked according to the similarity of their associated sequences with a test sequence. Each test subject was identified as the highest-ranking database subject.

In Ref. [5], each gait sequence was mapped to similarity plots which were derived by the self-similarities between each pair of images in the sequence. The feature vectors that were used for classification consisted of units of self-similarity of size one period. The feature vectors were scaled to a constant size in order to compensate for period differences. The scaled feature vectors were used for gait recognition using standard statistical pattern recognition techniques.

In Ref. [6], identification using Principal Component Analysis (PCA) was performed using patterns derived from the horizontal and vertical projections of silhouettes. The projections constitute periodic signals that were subsequently analyzed using the mathematical theory of symmetry groups. By studying the correlation between symmetry groups and gait viewing direction (with the help of a humanoid walking avatar), practical techniques were developed for classifying imperfect patterns.

In Refs. [7,8], Hidden Markov Models (HMM) [9] were used for the modelling of each gait sequence in the database set. The framework proposed in Refs. [7,8] is flexible since it is not dependent on a particular feature vector. For each test subject, the probability that its gait sequence was generated from one of the models associated with the database sequences was calculated. The subject corresponding to the model yielding the higher probability was considered to be identical to the test subject. The above approach yielded improved recognition performance when tested using modified sets of test sequences of the "Gait Challenge" database [4].

In Ref. [10], a gait recognition technique was proposed based on the extraction of key frames from silhouette sequences. Recognition was achieved using normalized correlation of silhouette templates and nearest neighbor classification. The use of shape as a biometric was investigated in Ref. [11]. The gait sequences were clustered into groups of similar frames. Test and reference cluster centers were compared and classification was performed using nearest neighbors.

In Ref. [12], silhouette extraction was achieved using a model-based method. These silhouettes had fewer noise pixels and missing parts. The silhouette sequences produced using this method were tested in conjunction with the model-based algorithm in Ref. [13]. The resultant system outperformed the baseline system in Ref. [4].

In Ref. [14], time varying signals were extracted from gait sequences using area-based metrics. The extracted signals were subsequently used as signatures for automatic gait recognition. It was observed that the best performance was achieved by combining information from different area masks. Moreover, other approaches were presented in Ref. [15] in which the motion of the knee and thigh during walking were modelled and processed. Such approaches, in general, require gait sequences of good quality in order to give good results.

Recently, an image analysis methodology that bears some similarity with the analysis technique proposed in this paper was presented [16]. In Ref. [16], a silhouette unwrapping technique was applied by calculating the distances of all contour pixels from the center of the silhouette. The collection of distances was used as a feature vector on which PCA was applied in order to reduce the dimensionality of the problem.

### 1.3. Present approach

In this work, we investigate a methodology for the efficient analysis of walking silhouettes and its applications in a gait recognition system. The proposed system bases its efficiency on a new method for the determination of the extent of similarity among the walking styles of subjects. This is combined with a technique for the preprocessing of gait sequences, prior to the application of the recognition

algorithm, in order to enhance the quality of the silhouettes [1]. By using the resultant system, the recognition performance on the Gait-challenge database will be seen to improve over the methods in Refs. [4,12,10]. Although our linear time normalization approach yields excellent results when used with raw silhouettes, we also propose a new transform which calculates a metric of the silhouette in angular slices of various orientations with respect to the center of the silhouette. In the experimental results section, this new feature will be compared with other features that were recently presented in the literature. In comparison to the method in Ref. [16], the new feature has several advantages: it is robust to segmentation errors, it does not require detection of contour pixel positions, and it can be computed easily.

The structure of the paper is as follows. Section 2 presents an overview of the proposed system. Section 3 presents the angular transform for gait analysis. The proposed gait recognition system using linear time normalization is presented in Section 4. This system is experimentally evaluated in Section 5 and finally, conclusions are drawn in Section 6.

## 2. System overview/problem statement

In this work, we assume that input to our system are *sequences of silhouettes* that have been extracted from video sequences depicting walking persons. A block diagram of the proposed system for the feature extraction and the partitioning of gait sequences into gait cycles is shown in Fig. 2. The preprocessing step involves operations on the raw silhouettes. These operations, which were recently presented in Ref. [1], have to do with the denoising of the sequences and their partitioning into gait cycles. The feature extraction process aims to convert each silhouette into a low-dimensional vector of coefficients. This can be achieved using our novel angular transform. Linear discriminant analysis (LDA) can additionally be used in order to reduce the dimensionality of the angular feature vector. It is worth noting that, in practice, the deployment of raw silhouettes instead of feature vectors yield excellent results.

The cycles of each gait sequence, determined during the preprocessing stage, are compared and their similarity is evaluated using the recognition/verification module which is based on a new distance calculation process. Our aim is to achieve as higher recognition/verification rates as possible.

## 3. Feature extraction

### 3.1. Angular gait analysis

One of the issues of major importance in a recognition system is feature extraction, i.e., the transition from the initial data space to a feature space that will make the recognition problem more tractable. In a gait recognition system, the feature extraction process should take into account the following prerequisites:

- *Easy scale compensation*. In practical recognition applications based on gait, humans may walk at different distances from the camera imposing a need to compensate the distance difference prior to the application of any gait recognition methodology. For this reason, it is important that the required scale compensation is performed without having to resize all frames captured by the monitoring device.
- *Translation invariance*. It is highly unlikely that in all frames of a gait sequence the observed subject will always be in the same position (in each frame). This would impose a need for alignment of frames with respect to the position of the human in the frame. A feature extraction process that yields translation invariant features would greatly facilitate and simplify the recognition task.
- *Robustness to segmentation errors*. Since, in a gait recognition system, the walking human has to be extracted from its background before processing, a segmentation process is required. This process, termed background subtraction, should be performed automatically without human intervention. The automatic background subtraction will always be imperfect. Therefore, it is crucial that the feature extraction process is not sensitive to artifacts incurred in the segmented frames due to the background subtraction process.
- *Low-computational complexity*. This is a requirement in many applications in which gait analysis and recognition or verification should be performed in real time. In particular, in several security applications such as airport security, border control, perimeter surveillance, only real-time recognition/verification would be acceptable. This imposes response-time limitations on gait recognition/verifiction systems which, therefore, have to be of low-computational complexity.

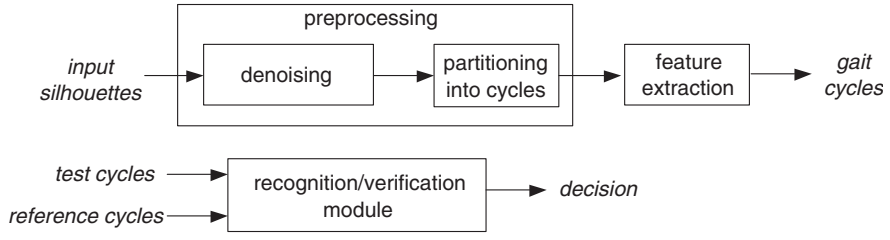


Fig. 2. Block diagram of the gait recognition/verification system.

Bearing the above issues in mind, we proceed in the formulation of a transform which satisfies all above prerequisites. As almost all recent approaches for gait recognition, we rely solely on binary silhouettes derived by background subtraction. Specifically, we assume that each gait sequence is composed of binary silhouettes $s[h, v]$, where $(h, v)$ is the pixel position. Let

$$s[h, v] = \begin{cases} 1 & \text{if } (h, v) \text{ belongs to the foreground,} \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

In order to apply our approach, first the area center $(h_c, v_c)$ of each silhouette is computed as

$$h_c = \frac{1}{N}\sum_{h,v} h \cdot s[h, v], \quad v_c = \frac{1}{N}\sum_{h,v} v \cdot s[h, v], \quad (2)$$

where $N$ is the number of foreground pixels, given by

$$N = \sum_{h,v} s[h, v]. \quad (3)$$

Once the center of the silhouette is calculated, we define a new coordinate system $x - y$, whose origin is at the center of the silhouette. We propose an angular transform:

$$\mathscr{A}(\theta) = \frac{1}{N_{\mathscr{F}_\theta}} \sum_{(x,y)\in\mathscr{F}_\theta} s[x, y]\sqrt{x^2 + y^2}, \quad (4)$$

where $\mathscr{F}_\theta$ is the set of foreground pixels in the circular sector $(\theta - \Delta\theta/2, \theta + \Delta\theta/2)$ and $N_{\mathscr{F}_\theta}$ is the number of pixels in $\mathscr{F}_\theta$. Actually, since $s[x, y] = 1$ for $(x, y) \in \mathscr{F}_\theta$, the term $s[x, y]$ could be omitted from Eq. (4). For each $\theta$, the transform coefficient $\mathscr{A}(\theta)$ expresses the average distance of foreground pixels (in the direction defined by $\theta$) from the center of the silhouette. In practice, since there is an infinite number of angles $\theta$, the angular transform is computed in slices of $\Delta\theta$. The angle step $\Delta\theta$ determines the level of detail of the transform. The transform is graphically illustrated in Fig. 4. As expected, the transform coefficients, which exhibit the larger variability through time correspond to 270° (the leg area). The coefficients corresponding to 90° (the head area) are of high magnitude but their variability through time is small.

The use of the proposed transform obviates the need to align the silhouettes to the center of the frame as most gait recognition methods demand. This is due to the fact that the proposed angular transform is inherently translation invariant, since it is always calculated with respect to the center of the silhouette. Moreover, the averaging that takes place in each direction of $\theta$ is practically implicitly equivalent with a low-pass filtering procedure which endows the resulting transform with robustness to segmentation errors. This is very important since, as mentioned earlier in this section, in practical cases the background subtraction process, that is usually performed automatically without human intervention, will always generate imperfect silhouettes.

The proposed transform is also very efficient in terms of computational complexity. It can be computed using two fast raster scans of the binary silhouette. In the first scan, the center of the silhouette is computed using Eq. (2). In the second scan, foreground (silhouette) pixels are classified in directionality bins according to their positions with respect to the center of the silhouette. A transform coefficient is calculated, using Eq. (4), as the average distance of pixels in the corresponding bin from the center of the silhouette.

### 3.2. Continuous-domain angular analysis

In order to gain insight into the angular transform, we can study an equivalent transform $\mathscr{A}_c(\theta)$ in the continuous domain defined as the integral of the distances of the foreground points from the center of the silhouette over the foreground area in a sector of angle $\Delta\theta$. The integral of the distances from the center of the silhouette is

$$V = \int_{\theta-\Delta\theta/2}^{\theta+\Delta\theta/2} \int_0^{r_\theta} r^2 \, \mathrm{d}r \, \mathrm{d}\theta = \Delta\theta \frac{r_\theta^3}{3}, \quad (5)$$

where $r_\theta$ is the distance of the boundary from the center of the image in the direction defined by $\theta$ (see Fig. 3). Similarly, the foreground area of the silhouette in a sector $\Delta\theta$ is

$$Q = \int_{\theta-\Delta\theta/2}^{\theta+\Delta\theta/2} \int_0^{r_\theta} r \, \mathrm{d}r \, \mathrm{d}\theta = \Delta\theta \frac{r_\theta^2}{2}. \quad (6)$$

Therefore, from Eqs. (5) and (6):

$$\mathscr{A}_c(\theta) = \frac{V}{Q} = \frac{2}{3} r_\theta. \quad (7)$$

From the above expression it is easily deduced that the proposed angular transform coefficients are proportional to the distances of the contour pixels from the center of the silhouette. Using our transform, the extraction of contour-related information is achieved without explicitly detecting any contour pixels and endows our scheme with the valuable additional advantage of robustness to spurious pixels that usually occur due to background subtraction. Therefore, our feature extraction approach is clearly differentiated from the approaches in Refs. [16,7] in which each element in a feature vector is dependent on contour pixel detection.

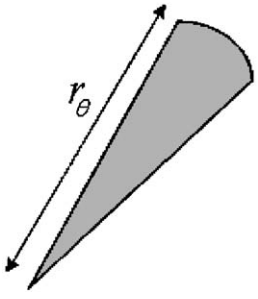


Fig. 3. Circular sector.

The proposed angular transform has a very convenient property which allows the scaling of the silhouettes in the transform domain by directly scaling the transform coefficients. From Eq. (7) it can be trivially deduced that if $\tilde{\mathscr{A}}(\theta)$ denotes the transform of a silhouette that was scaled by $\alpha$ in each dimension then

$$\tilde{\mathscr{A}}(\theta) = \alpha \mathscr{A}(\theta). \tag{8}$$

The above equation means that if the silhouette is scaled by $\alpha$ in each dimension, the transform $\tilde{\mathscr{A}}(\theta)$ of the scaled silhouette is equal to $\alpha \mathscr{A}(\theta)$. We use this property in order to achieve scale-invariance by scaling coefficients in each transformed silhouette so that the maximum coefficient in the silhouette is a constant value.

### 3.3. Feature matrix formulation

The gait recognition system we implemented is based on the angular silhouette representation described previously. All transform coefficients of a silhouette are ordered in a single gait vector of dimension $K = 360/\Delta\theta$ where the angle step $\Delta\theta$ denotes the width of the angle intervals in which the transform is computed. Simulations indicate that the recognition performance is largely independent of $\Delta\theta$ in the range 3–15°. The transform vector corresponding to a silhouette is of the form:

$$\mathbf{A} = [\mathscr{A}(\theta_0) \;\; \mathscr{A}(\theta_1) \;\; \ldots \;\; \mathscr{A}(\theta_{K-1})]. \tag{9}$$

Each transformed vector was scaled so that its maximum value is $M$. The algorithm is invariant to the exact value of $M$ as long as the same $M$ is used for all transformed sequences. For the results in the present paper $M = 255$. The scaling operation makes the proposed representation scale-invariant, a property which shall be very useful in most practical applications since humans may walk at different distances from the camera. The entire gait sequence is represented by a sequence of transform vectors of the above form. In practice a feature matrix $\mathbf{V}$ is formed as

$$\mathbf{V} = \begin{bmatrix} \mathscr{A}_0(\theta_0) & \mathscr{A}_1(\theta_0) & \cdots & \mathscr{A}_{L-1}(\theta_0) \\ \mathscr{A}_0(\theta_1) & \mathscr{A}_1(\theta_1) & \cdots & \mathscr{A}_{L-1}(\theta_1) \\ \vdots & \vdots & \ddots & \vdots \\ \mathscr{A}_0(\theta_{K-1}) & \mathscr{A}_1(\theta_{K-1}) & \cdots & \mathscr{A}_{L-1}(\theta_{K-1}) \end{bmatrix},$$

where $\mathscr{A}_l(\theta_k)$ denotes the $k$th transform coefficient in the $l$th frame and $L$ is the length of the sequence (in frames). A typical transform representation for a gait sequence is shown in Fig. 4(b) in which the values of the transform coefficients are plotted with respect to the frame index.

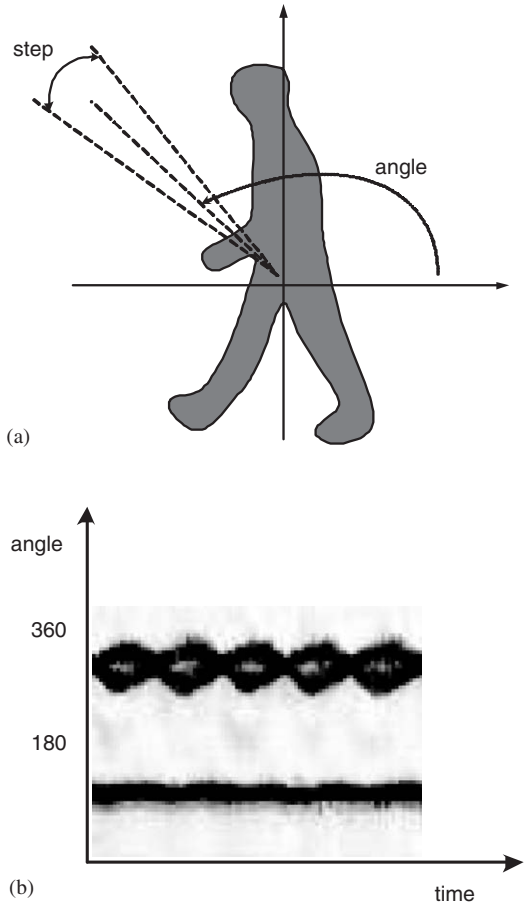


Fig. 4. (a) Graphical representation of the proposed transform. The *angle* defines the direction of the slice, whereas the angle *step* $\Delta\theta$ determines the width of the slice. (b) Typical transform representation of a silhouette sequence (the gray scale was inverted for displaying convenience). The legs area (270°) compacts most information.

## 4. Gait recognition

For the description of the recognition process, we adopt the terminology in Ref. [17], according to which the reference gait sequences will be termed *gallery* sequences, whereas the test gait sequences will be termed *probe* sequences. Moreover, for the sake of ease of reference, a short table is included (Table 1), explaining the notation used in the remainder of the section.

### 4.1. Linear time normalization

The gait cycles, which in this work are determined using the process in Ref. [2], will be of different lengths even within the same gait sequence. Therefore, the distance between cycles extracted from different gait sequences cannot in general be evaluated by direct comparison.

A reasonable approach to evaluating the distance between test and reference cycles of different lengths is to perform time normalization. This means that the cumulative distance is computed by summing the distances between each feature vector in a gallery (probe) cycle and its corresponding feature vector in the probe (gallery) cycle. In the

Table 1
Notation

| Symbol | Notation | Symbol | Notation |
|---|---|---|---|
| $i$ | Probe cycle index | $T_i$ | Number of frames in $i$th probe cycle |
| $j$ | Gallery cycle index | $R_j$ | Number of frames in $j$th gallery cycle |
| $\mathscr{C}_i^p$ | $i$th probe cycle | $d_{ij}$ | Distance between $\mathscr{C}_i^p$ and $\mathscr{C}_j^g$ |
| $\mathscr{C}_j^g$ | $j$th gallery cycle | $D_r^p$ | Probe subdistance |
| $I$ | Number of probe cycles | $D_r^g$ | Gallery subdistance |
| $J$ | Number of gallery cycles | $D_r$ | Total distance |

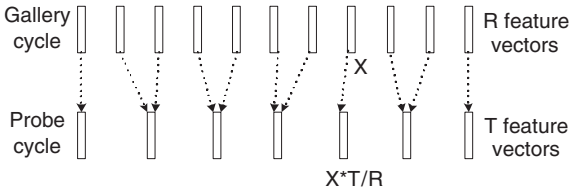


Fig. 5. Graphical representation of the linear time normalization process. The gallery cycle consists of $R$ frames whereas the probe cycle consists of $T$ frames (in general, $T \neq R$). Each gallery frame is compared with a probe frame which is determined by linearly compensating the length differences.

present work we use linear time normalization (henceforth termed LTN), i.e., the correspondence between frames in cycles of different length are computed using a linear rule. In general, time normalization can be performed linearly or nonlinearly. Nonlinear time normalization can be achieved using dynamic time warping [18]. However, *we found that in almost all cases the linear time normalization process yields better results and thus, it is more preferable than dynamic time warping*. This conclusion is in contrast to our intuitive expectation that is based on speech recognition paradigms, in which dynamic time warping was reported to be *much* more efficient than linear time normalization [18].

The LTN technique is depicted in Fig. 5. In practice, since the start and the end feature vectors in some cycles may not have been computed accurately, we allow for a slight deviation from the linear rule by examining two additional feature vectors in the vicinity of the nominal (linearly determined) feature vector. Therefore, the corresponding feature vector is defined as the one among the eligible vectors that yields the smallest distance.

The distance computation is performed between all cycles $\{\mathscr{C}_1^p, \mathscr{C}_2^p, \ldots, \mathscr{C}_I^p\}$ in the probe sequence and all cycles $\{\mathscr{C}_1^g, \mathscr{C}_2^g, \ldots, \mathscr{C}_J^g\}$ in the gallery sequence, where $I$ and $J$ denote the number of cycles in a probe and a gallery sequence, respectively. If $T_i$, $R_j$ denote the number of feature vectors in the $i$th probe cycle and the $j$th gallery cycle, $t$, $r$ are the feature vector indices in the probe and gallery cycles, respectively, and $\mathscr{T}_r$, $\mathscr{R}_t$ are the sets consisting of the nominal index (determined using the linear rule) and its two neighboring indices, then the distance between two cycles is defined as

$$d_{ij} = \frac{1}{2}\left(\sum_{t=1}^{T_i} \min_{r\in\mathscr{R}_t} d_e(\mathscr{A}_t, \mathscr{A}_r) + \sum_{r=1}^{R_j} \min_{t\in\mathscr{T}_r} d_e(\mathscr{A}_t, \mathscr{A}_r)\right). \tag{10}$$

The calculated distances

$$d_{ij} = LTN\left(\mathscr{C}_i^p, \mathscr{C}_j^g\right), \quad \text{for} \quad i=1,\ldots,I \text{ and } j=1,\ldots,J$$

form an $I \times J$ matrix $\mathscr{D}$ which will be used in the sequel for the determination of the distance between the probe sequence and the gallery sequence:

$$\mathscr{D} = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1J} \\ d_{21} & d_{22} & \cdots & d_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ d_{I1} & d_{I2} & \cdots & d_{IJ} \end{bmatrix}.$$

### 4.2. Final distance determination

Having formed the matrix with all distances between probe and gallery cycles, we need to find a way to calculate a distance $D_r$ between the probe sequence and the gallery sequence which will accurately represent the extent of similarity between the two gait sequences. A prerequisite which we may intuitively set in the formulation of an algorithm for determining $D_r$ is that the distance calculation process should be symmetric with respect to probe and gallery sequences, i.e., if the probe and gallery sequences were interchanged, the computed distance would be identical. Therefore, any process applied in a "probe-wise" manner should also be applied in a "gallery-wise" manner. For this reason, we treat each probe (gallery) cycle separately and evaluate its similarity with gallery (probe) cycles by taking the minimum distance of this probe (gallery) cycle to the gallery (probe) cycles with which it was compared. In practice, this means that we find the minima in each line and

each column of matrix $\mathscr{D}$, i.e.

$$d^p_{min,i} = \min_j \{d_{ij}\}, \quad \text{for } i = 1, \ldots, I,$$

$$d^g_{min,j} = \min_i \{d_{ij}\}, \quad \text{for } j = 1, \ldots, J.$$

The above distances form two sets:

$$\mathscr{S}_p = \{d^p_{min,1}, \ldots, d^p_{min,I}\} \quad \text{and}$$

$$\mathscr{S}_g = \{d^g_{min,1}, \ldots, d^g_{min,J}\}.$$

Before we proceed, we consider the following situation: *some cycles in the captured probe or gallery sequences may not be representative of a person's walking style. Distances which involve* (*one or two*) *such cycles may therefore be unreasonably small or large. Such distances should be excluded from consideration in our pursue for the representative distance* $D_r$. Since unavoidably, some cycles will not be representative, some minima in the sets $\mathscr{S}_p$ and $\mathscr{S}_g$ will not be reliable. Motivated by the above conclusion, we can take the mean or median of the distances in the sets $\mathscr{S}_p$ and $\mathscr{S}_g$:

$$D^p_r = mean/median\{d^p_{min,1}, \ldots, d^p_{min,I}\} \quad \text{and}$$

$$D^g_r = mean/median\{d^g_{min,1}, \ldots, d^g_{min,J}\}$$

and average them in order to calculate the final distance $D_r$ between the test and the reference sequence:

$$D_r = \tfrac{1}{2}\left(D^p_r + D^g_r\right). \tag{11}$$

This process is illustrated in Fig. 6.

The rationale behind choosing to average the two median (or mean) values instead of taking the minimum of them is explained below. Assume that the probe and gallery cycles correspond to two different subjects and that there is a particular gallery cycle $\mathscr{C}^g_z$ which resembles the probe cycles. In this case, the minimum distance between $\mathscr{C}^g_z$ and $\mathscr{C}^p_1, \mathscr{C}^p_2, \ldots, \mathscr{C}^p_I$ would be small. Most of these sub-distances would be included in $\mathscr{S}_p$ and consequently, the median (or mean) operator would yield a small distance. If the smallest of the two was taken as the eventual distance, then the resulting distance $D_r$ would be mistakenly small. Therefore, it is important that sub-distances in both directions contribute to the calculation of the final distance. As will be seen in the experimental results section, this methodology yields good results in practice.

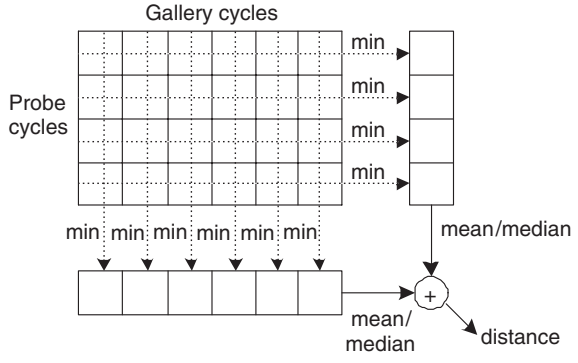


Fig. 6. Graphical representation of the proposed distance calculation process using a matrix comprising of distances among all cycles in a probe sequence against all cycles in a gallery sequence.

### 4.3. Recognition and verification

For a given test gait sequence, its distance from all the sequences in a reference database is calculated in order to determine a matching gait sequence. Since a smaller $D$ means a closer match, the best match of a probe sequence in a set of sequences is identified as the sequence with which the above distance is minimum.

For verification purposes, the claimed identity of a subject is verified if the distance between the subject sequence and the claimed gallery subject sequence is below a predetermined threshold. This threshold controls the sensitivity of verification and is chosen according to the specific application.

## 5. Experimental results

### 5.1. Database description

The methodologies of the previous sections were tested on USF's Gait Challenge database of human gait sequences. The sequences in this database were captured under different conditions. The gallery (reference) set of gait sequences was used as the system database and the probe (test) sets A–G were considered to contain sequences of unknown subjects who are to be recognized by comparison of their gait sequences to the sequences in the gallery set.[1] The possible differences between the capturing conditions of the probe sequences and the gallery sequences are with respect to walking surface (cement/grass), shoe type (type A/B), and view-angle (left/right). The gallery set contains 71 gait sequences of individuals walking on Grass, wearing type-A shoes, and was captured using the Right camera. Table 2 summarizes the exact conditions for the recording of the probe sequences, in where C, G, A, B, L, R, stand for Cement, Grass, shoe type A, shoe type B, Left view, and Right view, respectively. The number of sequences is each set is in square brackets.

It should be noted here that the gallery contains only one sequence/subject. This resembles the particular, but very likely, situation in which only one gait sequence/subject might be available in a reference database. If more sequences were available for a subject, more distances given by Eq. (11) would have to be calculated.

[1] We used the May 2001 silhouettes. We removed from probe sets D, E, F, G the subjects that were not in the Gallery.

Table 2
Differences in capturing conditions between probe sets A–G and the gallery

| Probe set | Difference | Probe set | Difference |
|---|---|---|---|
| A (GAL) [71] | View | E (CBR) [42] | Surface, shoe |
| B (GBR) [41] | Shoe | F (CAL) [66] | Surface, view |
| C (GBL) [41] | Shoe, view | G (CBL) [42] | Surface, shoe, view |
| D (CAR) [66] | Surface | | |

The conditions under which each probe sequence was recorded are summarized in brackets. The number of subjects in each set is reported in squared brackets.

### 5.2. Recognition performance

For evaluating the recognition performance of our schemes, we report Cumulative Match Scores (as in Ref. [17]) ranks at 1 and 5. Rank 1 results report the percentage of the subjects in a probe set that were identified exactly. Rank 5 results report the percentage of probe subjects whose true match in the gallery set was in the top 5 matches.

Initially, we examined the effect of $\Delta\theta$ variation on the performance of the algorithm that uses the angular feature. Cumulative match scores derived for several $\Delta\theta$ values are reported for rank-1 and rank-5 in Figs. 7(a) and (b), respectively. As seen, the recognition performance is not critically dependent on $\Delta\theta$ variations. However, using a larger angle step is usually beneficial for the rank-5 scores at the cost of a lower rank-1 score and vice-versa. It seems that the less detailed transform resulted by larger angle step is more appropriate for approximate (rank-5) recognition but less appropriate for exact (rank-1) recognition. In this work, we use $\Delta\theta = 5$ for all our experiments.

In the sequel, we report results of the angular feature in comparison to commonly used features such as the width of silhouette and a feature derived from the concatenation of the horizontal and vertical projections of the silhouette. Results in terms of recognition rates at ranks 1 and 5 are reported in Table 3. As seen, the angular feature clearly outperforms the other two features, proving its appropriateness for gait representation.

Cumulative Match Scores are presented in Fig. 8. We report results for our linear time normalization (LTN) algorithm using the silhouette feature (LTN-S) and the feature based on the 30 most important coefficients (determined using LDA) of the angular representation (LTN-A). Results are reported in Table 4 and are compared to USF's baseline algorithm [4], the methods by the Massachusetts Institute of Technology [12] and Carnegie Mellon University [10].[2] These methods are in general more complex than ours. Specifically, the method in Ref. [12] employs model-based silhouette and feature extraction, whereas the method

[2] Results for the method in Ref. [10] on the Gait Challenge database were reported in Ref. [11].

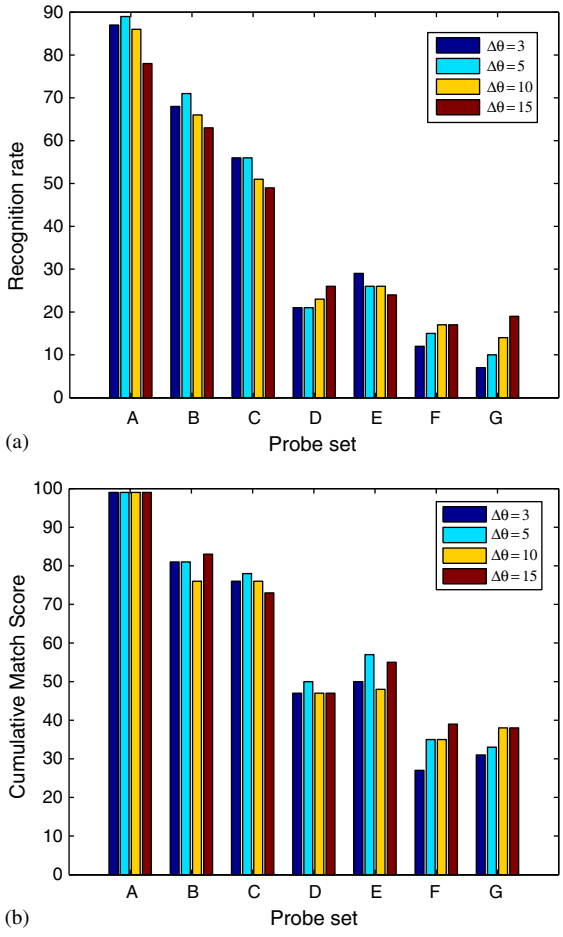


Fig. 7. The effect of the variation of $\Delta\theta$ on the performance of the proposed gait recognition methodology: (a) rank-1; (b) rank-5.

Table 3
Comparison of several features to the proposed angular feature

| | Rank 1 | | | Rank 5 | | |
|---|---|---|---|---|---|---|
| | Angular | Width | Projections | Angular | Width | Projections |
| A | 89 | 69 | 83 | 99 | 85 | 96 |
| B | 71 | 71 | 78 | 81 | 78 | 83 |
| C | 56 | 44 | 71 | 78 | 68 | 78 |
| D | 21 | 17 | 12 | 50 | 36 | 32 |
| E | 26 | 14 | 5 | 57 | 33 | 31 |
| F | 15 | 14 | 9 | 35 | 27 | 24 |
| G | 10 | 7 | 5 | 33 | 31 | 26 |

The probability of identification (in percent) at ranks 1 and 5 is reported.

in Ref. [10] compares silhouette templates using exhaustive correlation, at all pixel displacements, between the two templates. As seen in Table 4, the proposed method using LTN is able to capture more gait information than the baseline algorithm since the results in most cases are better than the results reported in Ref. [4]. From Table 4, we observe that our silhouette-based method clearly outperforms the methods in Refs. [12,10] by approximately

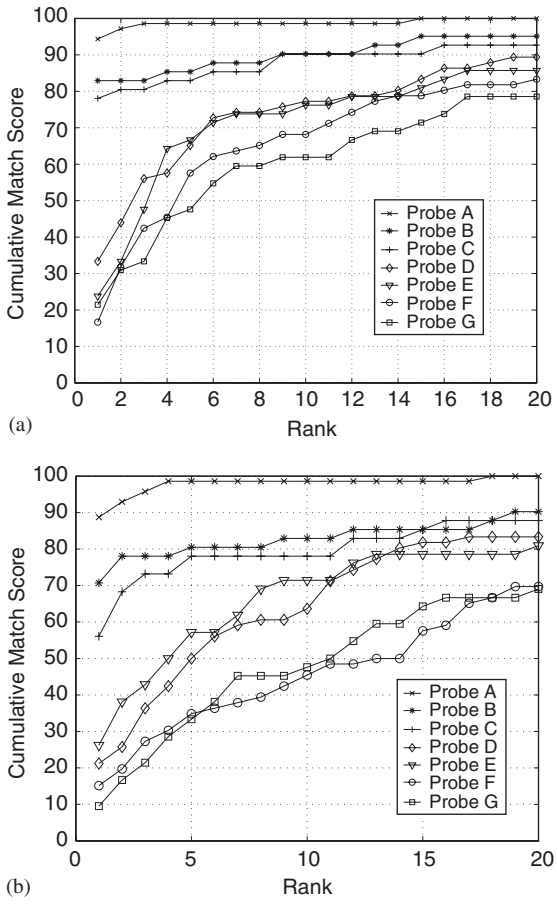


Fig. 8. Cumulative Match Scores for probes A–G using the proposed method. (a) silhouette feature; (b) angular feature.

10% and 8%, respectively (in terms of rank-1 performance). Considering that our approach is quite simple, we regard these gains as largely satisfactory, although there is still room for improvement, mainly in the direction of boosting the performance of the even simpler angular-based system so that it reaches the performance of the silhouette-based system.

Since both LTN-S and the method in Ref. [4] use the same feature, the gains achieved using LTN-S come from the denoising process [1] which allows more accurate calculation of structural similarities, and the LTN process which exploits the discrimination power of dynamic information, i.e., the information about the succession of stances during walking. It should be noted here that the LTN achieves reliable matching even when the test and the reference subjects walk at different speeds. In such a case the baseline USF algorithm would not yield reliable distance calculations.

### 5.3. Verification performance

The proposed system was also evaluated in terms of verification performance. The most widely used method for this task is to present rate operating characteristic curves (ROC). In an access control scenario, this means calculating the probability of positive recognition of an authorized subject versus the probability of granting access to an unauthorized subject. For calculating the above probabilities, we vary the distance that determines which subjects are granted access and which are not. For this calculation we use normalized distances (as in Ref. [19]) or unnormalized distances. Using normalized distances is in general more suitable for verification purposes.

Verification results are tabulated in Table 5. The ROC's in their entirety are also displayed in Fig. 9. As seen, the proposed system based on the silhouette feature generally outperforms the baseline method. Despite the good performance of both schemes, at present, it appears that it might not be used on its own for the purpose of authentication without the use of additional biometrics or additional access control mechanisms. However, for interpreting these results, we should also take into consideration the fact that the database used for evaluating verification performance was captured outdoors and is very noisy. In a practical access control system, the capturing environment could be controlled, e.g., with uniform background, and multiple cameras could be used for generating high-quality gait sequences. In such a

Table 4
Comparison to several algorithms using the Gait Challenge database

| Probe | Rank 1 | | | | | Rank 5 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | LTN-S | LTN-A | MIT [12] | CMU [10] | USF [4] | LTN-S | LTN-A | MIT [12] | CMU [10] | USF [4] |
| A | 94 | 89 | 87 | 87 | 79 | 99 | 99 | 96 | 100 | 96 |
| B | 83 | 71 | 83 | 81 | 66 | 85 | 81 | 90 | 90 | 81 |
| C | 78 | 56 | 66 | 66 | 56 | 83 | 78 | 88 | 93 | 76 |
| D | 33 | 21 | 25 | 21 | 29 | 65 | 50 | 50 | 59 | 61 |
| E | 24 | 26 | 23 | 19 | 24 | 67 | 57 | 48 | 50 | 55 |
| F | 17 | 15 | 20 | 27 | 30 | 58 | 35 | 49 | 53 | 46 |
| G | 21 | 10 | 19 | 23 | 10 | 48 | 33 | 50 | 43 | 33 |

The probability of identification (in percent) at ranks 1 and 5 is reported.

Table 5
Verification performance of the proposed scheme using the silhouette feature

| Probe | Unnormalized | | | | $z$-Normalized | | | | USF (unnormalized) |
|---|---|---|---|---|---|---|---|---|---|
| | 1% | 5% | 10% | 20% | 1% | 5% | 10% | 20% | 10% |
| A | 79 | 90 | 97 | 98 | 99 | 99 | 99 | 99 | 86 |
| B | 63 | 78 | 83 | 87 | 83 | 88 | 89 | 95 | 76 |
| C | 41 | 66 | 77 | 85 | 76 | 80 | 89 | 93 | 59 |
| D | 11 | 32 | 39 | 65 | 27 | 60 | 75 | 83 | 42 |
| E | 7 | 23 | 42 | 71 | 22 | 57 | 74 | 81 | 52 |
| F | 8 | 27 | 36 | 55 | 17 | 44 | 62 | 80 | 41 |
| G | 8 | 22 | 39 | 58 | 20 | 40 | 62 | 74 | 36 |

Unnormalized and normalized verification rates are reported for several false alarm probabilities. The (unnormalized) verification performance of USF's Baseline algorithm [4] is included for 10% false alarm probability.

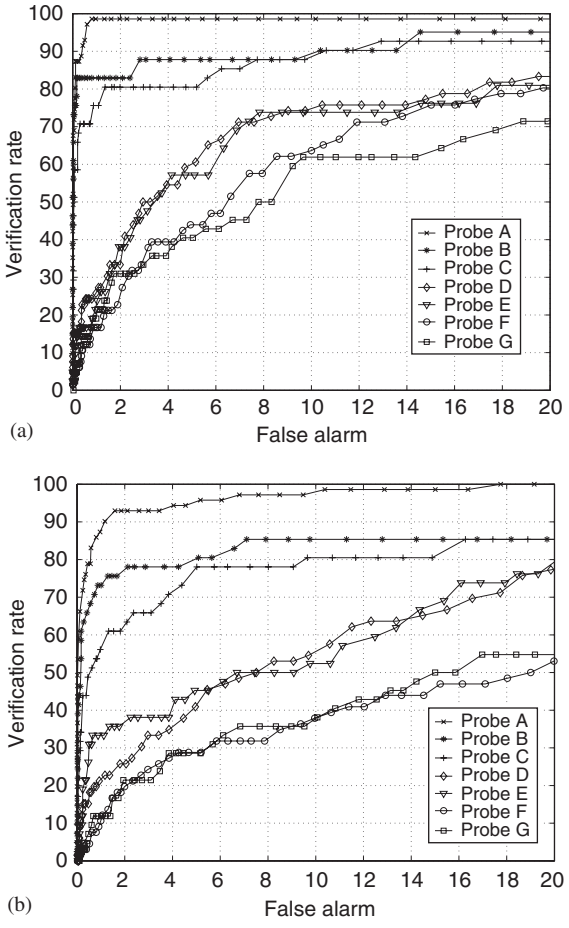


Fig. 9. Rate operating characteristics for probe sets A–G using the proposed methods. Normalized distances were used: (a) silhouette feature; (b) angular feature.

controlled environment, we expect that the verification performance would be significantly better allowing the deployment of gait-assisted verification in practical cases.

## 6. Conclusions

A new method was presented for gait-assisted recognition from sequences of binary silhouettes. The new method is based on linear time normalization for the efficient matching between probe and reference sequences. Although the proposed method is directly applicable on binary silhouettes, a transform representation of silhouettes was also proposed which has many attractive features that make it very suitable for deployment in gait analysis and recognition applications. The sequence of the silhouettes or the sequence of the angular feature vectors were used for identity recognition or verification based on the matching of linearly time-normalized gait walking cycles. By using the best-performing of the proposed methodologies, improvements of 8–20% in recognition and verification performance were seen in comparison to other known methodologies on the "Gait Challenge" database.

## References

[1] N.V. Boulgouris, K.N. Plataniotis, D. Hatzinakos, An angular transform of gait sequences for gait assisted recognition, in: IEEE International Conference on Image Processing, Singapore, 2004.
[2] N.V. Boulgouris, K.N. Plataniotis, D. Hatzinakos, Gait recognition using dynamic time warping, in: IEEE International Symposium on Multimedia Signal Processing, Siena, Italy, 2004.
[3] S.A. Niyogi, E.H. Adelson, Analyzing and recognizing walking figures in XYT, in: Proceedings of Computer Vision and Pattern Recognition, 1994.
[4] P.J. Phillips, S. Sarkar, I. Robledo, P. Grother, K.W. Bowyer, The gait identification challenge problem: data sets and baseline algorithm, in: International Conference on Pattern Recognition, 2002.
[5] C. BenAbdelkader, R. Cutler, L. Davis, Motion-based recognition of people in eigengait space, in: IEEE International Conference on Automatic Face and Gesture Recognition, 2002.
[6] Y. Liu, R. Collins, Y. Tsin, Gait sequence analysis using frieze patterns, in: European Conference on Computer Vision, Copenhagen, 2002.
[7] A. Sundaresan, A.K.R. Chowdhury, R. Chellappa, A hidden Markov model based framework for recognition of humans from gait sequences, in: Proceedings of the ICIP 2003, Barcelona, Spain, 2003.
[8] A. Kale, Algorithms for gait-based human identification from a monocular video sequence, Ph.D. Dissertation, University of Maryland, College Park, 2003.
[9] L.R. Rabiner, A tutorial on hidden Markov models and selected applications on speech recognition, Proc. IEEE 77 (2) (1989) 257–285.

[10] R.T. Collins, R. Gross, J. Shi, Silhouette-based human identification from body shape and gait, in: IEEE Conference on Automatic Face and Gesture Recognition, 2002.
[11] D. Tolliver, R.T. Collins, Gait shape estimation for identification, in: Fourth International Conference on Audio and Video-Based Biometric Person Authentication, Guilford, UK, 2003.
[12] L. Lee, G. Dalley, K. Tieu, Learning pedestrian models for silhouette refinement, in: International Conference on Computer Vision, 2003.
[13] L. Lee, W.E.L. Grimson, Gait analysis for recognition and classification, in: Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition, 2002.
[14] J.P. Foster, M.S. Nixon, A. Prugel-Bennett, Automatic gait recognition using area-based metrics, Pattern Recognition Lett. 24 (2003) 2489–2497.
[15] D. Cunado, M.S. Nixon, J.N. Carter, Automatic extraction and description of human gait models for recognition purposes, Comput. Vision Image Understanding 90 (1) (2003) 1–14.
[16] L. Wang, T. Tan, H. Ning, W. Hu, Silhouette analysis-based gait recognition for human identification, IEEE Trans. Pattern Anal. Mach. Intell. 25 (12) (2003) 1505–1518.
[17] P.J. Phillips, H. Moon, S. Rizvi, P. Raus, The feret evaluation methodology for face recognition algorithms, IEEE Trans. Pattern Anal. Mach. Intell. 22 (10) (2000) 1090–1104.
[18] H. Sakoe, S. Chiba, Dynamic programming algorithm optimization for spoken word recognition, IEEE Trans. Acoust. Speech Signal Process. 26 (1) (1978) 43–49.
[19] P. Grother, Face recognition vendor test 2002, Supplemental Report, NISTIR 7083, available from http://www.frvt.org