

Self-Calibrating View-Invariant Gait Biometrics

Michela Goffredo, *Member, IEEE*, Imed Bouchrika, *Member, IEEE*, John N. Carter, *Member, IEEE*, and Mark S. Nixon, *Associate Member, IEEE*

Abstract—We present a new method for viewpoint independent gait biometrics. The system relies on a single camera, does not require camera calibration, and works with a wide range of camera views. This is achieved by a formulation where the gait is self-calibrating. These properties make the proposed method particularly suitable for identification by gait, where the advantages of completely unobtrusiveness, remoteness, and covertness of the biometric system preclude the availability of camera information and specific walking directions. The approach has been assessed for feature extraction and recognition capabilities on the SOTON gait database and then evaluated on a multiview database to establish recognition capability with respect to view invariance. Moreover, tests on the multiview CASIA-B database, composed of more than 2270 video sequences with 65 different subjects walking freely along different walking directions, have been performed. The obtained results show that human identification by gait can be achieved without any knowledge of internal or external camera parameters with a mean correct classification rate of 73.6% across all views using purely dynamic gait features. The performance of the proposed method is particularly encouraging for application in surveillance scenarios.

Index Terms—Gait biometrics, human identification, view invariant.

I. INTRODUCTION

GAIT BIOMETRICS, which concerns recognizing individuals by the way they walk, is a particularly challenging research area. The potential for personal identification is supported by a rich literature, including medical and psychological studies [1], [2]. The completely unobtrusiveness without any subject cooperation or contact for data acquisition makes gait particularly attractive for identification purposes. Gait recognition techniques at the state of the art can be divided into 3-D and 2-D approaches [3]. In the first group, identification relies on parameters extracted from the 3-D limb movement. These methods use a large number of digital cameras, and the 3-D reconstruction is achieved after a camera calibration process. On the other hand, the 2-D gait biometric approaches extract explicit features describing gait by means of human body models [4], [5] or silhouette shape [6], [7]. A

Manuscript received March 6, 2009; revised July 23, 2009. This paper was recommended by Associate Editor G. Zhao.

M. Goffredo is with the Department of Applied Electronics, University Roma Tre, 00145 Rome, Italy, and also with the School of Electronics and Computer Science, University of Southampton, SO17 1BJ Southampton, U.K. (e-mail: mg2@ecs.soton.ac.uk).

I. Bouchrika, J. N. Carter, and M. S. Nixon are with the School of Electronics and Computer Science, University of Southampton, SO17 1BJ Southampton, U.K. (e-mail: ib04r@ecs.soton.ac.uk; jnc@ecs.soton.ac.uk; msn@ecs.soton.ac.uk).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSMCB.2009.2031091

rich variety of data has been collected for the evaluation of 2-D gait biometrics. The widely used and compared databases on gait recognition include the University of South Florida [8], Carnegie Mellon University (CMU) [9], and the University of Southampton [10] data. The majority of methods and databases found in the literature use a single camera positioned with specific orientations with respect to the subject's walking direction (generally capturing the walk from the lateral view), and a large number of papers describing gait recognition have been published [3], [11].

In surveillance scenarios, we need a system that operates in an unconstrained environment where, maybe, there is no information regarding the camera [12] and where the subject walks freely. Gait recognition has contributed to evidence for convictions in criminal cases like the case of the murderer of Swedish Foreign Minister Anna Lindh [15], a bank robber in Noerager (Denmark) [13], and a burglar in Lancashire (U.K.) [14]. Lynnerup *et al.* [15] affirmed the usefulness of gait analysis in forensics. They were able to identify the two bank robbers by matching surveillance images with images of the suspects, and this evidence was later used to convict two suspects.

The aim of this paper is to present a novel 2-D view-independent gait biometrics approach which does not need camera calibration or preknowledge of subject pose, so that it can be used for gait identification in scenarios where only a single remote camera view is available.

In our approach, information on the human silhouette and human body anthropometric proportions [16] is used to estimate the lower limbs' pose in the image reference system, with low computational cost. After a markerless motion estimation, the obtained angles' trends are corrected by the viewpoint independent gait reconstruction algorithm that allows reconstruction of the limbs' pose in the sagittal plane for identification purposes.

The remainder of this paper is organized as follows. Section II summarizes the previous approaches on 2-D view-invariant gait biometrics found in literature and the publically available databases. The theoretical description of the proposed view-invariant approach is presented in Section III. Section IV introduces the experimental results and analysis. Finally, this paper is concluded in Section V with discussion and future developments.

II. RELATED WORK

Much research has been done into identifying subjects by how they walk from 2-D video data with publications dating as far back as 1994 [17].

Typically, gait analysis techniques can be described as either model based or statistical appearance based. Model-based

analysis usually involves fitting a model representing various aspects of the human anatomy to the video data and then extracting and analyzing its parameters. Conversely, appearance-based statistical analysis regards a subject's silhouette shape or the statistical distribution of pixels in the silhouettes and how it varies over time.

In this section, we describe the publically available multiview gait data sets and the 2-D approaches that use geometry considerations to derive a viewpoint-invariant gait signature.

A. Multiview Data Sets

Two-dimensional view-invariant gait identification is an emerging research field which makes gait biometrics closer to application in real scenarios [18]. While many data have been collected for 2-D lateral view gait biometrics [3], few data sets image walking people from different viewpoints.

In 2000, the University of Southampton published the small SOTON gait database, where ten subjects have been acquired simultaneously from three views with different covariates [10].

A relative large number of papers have used the small SOTON gait database and its extended successor, testing methods only on the lateral view and reaching 100% correct recognition rate. The majority are statistical approaches, like the velocity-moment-based method of Shutler *et al.* [19], the symmetry operator of Hayfron-Acquah *et al.* [20], the gait mask of Foster *et al.* [21], the statistical silhouette's shape analysis of Wang *et al.* [22], or the 2-D polar-plane silhouette description of Chen and Gao [23].

In 2001, the Center for Biometrics and Security Research (CBSR) of the Chinese Academy of Sciences (CASIA) introduced a multiview data set, called NLPR (successively named as CASIA-A) composed of 20 subjects and three different walking directions [24]. A large number of papers have been tested on this database, obtaining 92.5% recognition rate on the lateral view. Most of them are statistical based, including the approach of Yang *et al.* [25], the mass vector method of Hong *et al.* [26], the fusion methodology of Nizami *et al.* [27], the 2-D polar-plane silhouette description of Chen and Gao [23], and the silhouette spatiotemporal spectrum of Lam *et al.* [28].

Recently, CASIA published a large multiview gait database called CASIA-B [24], where 124 subjects have been captured from 11 synchronized cameras at different viewpoints. The data present different covariates.

The polar representation of a walking subject proposed by Chen and Gao [23] was tested only on the lateral view, while Bashir *et al.* [29] and Yu *et al.* [30] tested their view-independent approach by comparing different views.

B. View-Invariant Literature

Two-dimensional view-independent gait identification is a recent research area, and the approaches found in literature can be broadly divided as pose-based and pose-free methods. The pose-free approaches extract some gait parameters which are independent from the camera point of view [7], [22], [23], [26], [31], [32]. On the other hand, the pose-based method aims to synthesize the lateral view of the human body from an arbitrary viewpoint [33]–[35].

One of the first pose-free approaches was presented by Johnson and Bobick [32], who developed a multiview gait recognition method using static body parameters where human body limbs are labeled by analyzing the binary silhouette of the subject. Head, pelvis, and feet positions were extracted, following the geometrical proportions of human body segments. Experimental tests regarded 18 subjects walking in front of a single camera positioned at 45° and 90° with respect to the subject. The side-view data were captured at two different distances, 3.9 and 8.3 m from the camera. The results reported both scale and view-angle adjustments, and a recognition rate of 94% was achieved.

In 2002, BenAbdelkader *et al.* proposed a different approach, where the moving person is detected and tracked and an image template corresponding to the person's motion blob is extracted in each frame [31]. Subsequently, a self-similarity plot from the obtained sequence of templates has been computed. Experimental tests on outdoor sequences of 44 people with four sequences of each taken on two different days achieved a classification rate of 77%. The method has also been tested on indoor sequences of seven people walking on a treadmill, taken from eight different viewpoints (from 0° to 120°) and on seven different days. A classification rate of 78% was obtained for near-frontoparallel views and 65% on average over all views.

A statistical shape analysis is the solution proposed by Wang *et al.* [7], where the temporal changes of the detected silhouettes are represented by a 2-D vector, composed of the edge point distance to the centroid position over time. The gait signature is then obtained via a supervised pattern classification technique based on the full Procrustes distance measure. The method has been tested on the CASIA-A gait database. The same authors also applied principal component analysis on the 2-D silhouette representation [22], reaching a recognition rate of 84.1% on average over all views of CASIA-A data set.

More recently, Hong *et al.* [26] introduced a new representation for human gait recognition, called the mass vector and defined as the number of pixels with a nonzero value in a given row of the silhouette. Sequences of temporally ordered mass vectors have been used for gait representation, and tests of the CASIA-A gait database have given a correct classification rate (CCR) exceeding 95%.

The pose-based methods, which generate the lateral view from data acquired at different arbitrary views, are the most recent approaches to 2-D view-independent gait biometrics. This choice is justified by the fact that the lateral view has proven recognition capability with many approaches [3], [36]–[42] and that the pose-free approach works on a small number of camera positions.

In 2003, a group at the University of Maryland developed a gait recognition algorithm showing that, if a person is far enough from a single camera, it is possible to synthesize the lateral view from any other arbitrary view by knowing the camera calibration parameters [33]. The method has been tested on 12 people walking along straight lines at different camera views. Considering a gallery of people walking at lateral view, the video sequences where people walk at arbitrary angles were chosen as probes, and the receiver operating characteristic was computed for each view. More recently, further experimental

tests have been done by the same group [43], and the recognition performance has been extracted on the National Institute of Standards and Technology and CMU databases. A comparison with the nonnormalized images has been included, and the mean recognition rate has been rising from 35% to 65%.

The biometrics research group of the University of Southampton has focused attention on 2-D view-invariant gait recognition from 1999 [44], where a trajectory-invariant gait signature was presented. The method of Carter and Nixon corrects the variations in gait data by knowing the walking trajectory and modeling the thigh as a simple pendulum. The approach was then reformulated by Spencer and Carter [45] to provide a pose invariant biometric signature which did not require knowledge of the subject's trajectory. Results on synthesized data showed that simple pose correction for geometric targets generalizes well for objects on the optical axis. More recently, these techniques have been refined and applied on subjects wearing reflective markers with success observed from six different point views [35], [46].

III. PROPOSED METHOD: THEORY

Since the majority of papers found in literature rely on a small range of camera views or have been tested on few subjects, the aim of this paper is to introduce a new view-independent approach for gait biometrics and evaluate its performance on several different databases and conditions.

The proposed approach for view-independent gait biometrics is pose based and aims at human identification by projecting the limbs' motion into the lateral plane. It is based on three consecutive stages: markerless joint estimation from the image sequence, viewpoint rectification, and biometric identification.

A. Markerless Joints' Position Estimation

Let $\mathbf{S}(\mathbf{x}, \mathbf{y}, t_\varphi)$ be the red, green, and blue frame (of size $R \times C$ pixels) at time $\mathbf{T} = [t_1, t_2, \dots, t_\varphi, \dots, t_F]$, where (x_1, y_1) is the top-left corner of the image. By applying a background subtraction method based on the threshold of the three components of the color space YUV , the binary image $\bar{\mathbf{S}}(\mathbf{x}, \mathbf{y}, t_\varphi)$ is extracted. The pixels $(\mathbf{x}_{\text{sil}}, \mathbf{y}_{\text{sil}})$ containing the human silhouette are selected as the object with maximum area [47]. The lower limb pose estimation algorithm is based on the proportions of the human body segments, following the medical results of anatomical studies [16]

$$\begin{aligned} y'_{\text{hip}} &= \min(\mathbf{y}_{\text{sil}}) + 0.50 \cdot H \\ y'_{\text{knee}} &= \min(\mathbf{y}_{\text{sil}}) + 0.75 \cdot H \\ y'_{\text{ankle}} &= \min(\mathbf{y}_{\text{sil}}) + 0.90 \cdot H \end{aligned} \quad (1)$$

where H is the silhouette's height.

Subsequently, the shins and thighs have been analyzed separately. Let $\hat{\mathbf{X}}$ be the set of horizontal coordinates where $\bar{\mathbf{S}}(\hat{\mathbf{X}}, y'_{\text{knee}}, \hat{t}_\delta) = 1$; then, the subset of frames

$$\hat{\mathbf{T}} = [\hat{t}_1, \hat{t}_2, \dots, \hat{t}_\delta, \dots, \hat{t}_{D \leq F}] \quad (2)$$

where the legs that do not overlap are those where $\hat{\mathbf{X}}$ forms two single connected components larger than $0.02H$ pixels.

Consequently, for each frame \hat{t}_δ , the shin extraction algorithm, based on the linear approximation of the skeleton of the portion of image that includes the lower leg, is applied on $\bar{\mathbf{S}}(\mathbf{x}, \mathbf{y}, \hat{t}_\delta)$. The subimages containing the shins (labeled as $\mathbf{S}_{\text{shin1}}$ and $\mathbf{S}_{\text{shin2}}$) are extracted from

$$\mathbf{S}_{\text{lower leg}}(\mathbf{x}, \mathbf{y}, \hat{t}_\delta) = \begin{cases} \bar{\mathbf{S}}(\mathbf{x}, y_s, \hat{t}_\delta), & \text{if } y_s \in \mathbf{y}_{\text{shin}} \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where the vertical coordinates belonging to the two shins $\mathbf{y}_{\text{shin}} = [y_1, y_2, \dots, y_s, \dots, y_S]$ are defined in the following way:

$$\mathbf{y}_{\text{shin}} = [y'_{\text{knee}}, y'_{\text{knee}} + 1, \dots, \min(y'_{\text{ankle}}, y_{\text{overlap}} - 1)] \quad (4)$$

which includes the possibility of feet overlapping with y_{overlap} : vertical coordinate where the set of correspondent horizontal coordinates $\mathbf{x}_{\text{overlap}}$ belong to a single object larger than $0.02H$ pixels.

Therefore, the left and right shins are defined by the pixels with coordinates

$$\begin{cases} \mathbf{x}_{\text{shin}\ell}^{\text{skel}}(\hat{t}_\delta) = [x_{1\ell}, x_{2\ell}, \dots, x_{s\ell}, \dots, x_{S\ell}] \\ \mathbf{y}_{\text{shin}\ell}^{\text{skel}}(\hat{t}_\delta) = \mathbf{y}_{\text{shin}} \end{cases}, \quad \ell = \{1, 2\} \quad (5)$$

where

$$x_{s\ell} = \sum_{j=1}^C x_j \cdot \mathbf{S}_{\text{shin}\ell}(x_j, y_s, \hat{t}_\delta) / \sum_{j=1}^C \mathbf{S}_{\text{shin}\ell}(x_j, y_s, \hat{t}_\delta). \quad (6)$$

Then, the shins can be linearly approximated by the first-order polynomial with coefficients

$$\mathbf{p}_\ell(\mathbf{x}_{\text{shin}\ell}^{\text{skel}}, \hat{t}_\delta) = p_{\ell 0}(\hat{t}_\delta) + p_{\ell 1}(\hat{t}_\delta) \cdot \mathbf{x}_{\text{shin}\ell}^{\text{skel}}(\hat{t}_\delta). \quad (7)$$

Therefore, the angle between the shins and the vertical axis (called knee angle) at frame \hat{t}_δ is $\alpha_\ell(\hat{t}_\delta) = \pi - \arctan[p_{\ell 1}(\hat{t}_\delta)]$. The procedure is applied for each frame in the subset $\hat{\mathbf{T}} = [\hat{t}_1, \hat{t}_2, \dots, \hat{t}_\delta, \dots, \hat{t}_D]$.

In order to also be able to estimate α_1 and α_2 when the legs overlap, the extracted shins are labeled as "left" and "right" according to the following rationale without losing generality: The subject is walking for at least two gait cycles; in t_1 , the shin labeled as shin1 belongs to the right leg, and the labels swap after the legs overlap.

Subsequently, a third-order polynomial interpolation of α_1 and α_2 has been applied, and the knee angles are achieved during the whole video sequence. The choice of this polynomial has been determined experimentally.

The upper leg orientation is extracted at every frame $\mathbf{T} = [t_1, t_2, \dots, t_\varphi, \dots, t_F]$ with a coarse to fine hip estimation procedure where, at first, the hip position is achieved with

$$\begin{cases} x'_{\text{hip}\ell} = \frac{1}{P} \cdot \sum_{j=1}^P \tilde{x}_j + (2\ell - 3) \cdot H \cdot \mu \cdot 10^{-3} \\ y'_{\text{hip}\ell} = y'_{\text{hip}} \cdot (2\ell - 3) \cdot \left(\frac{\tilde{x}_P - \tilde{x}_1}{2} \right) \cdot \sin(0.3 \cdot \mu) \end{cases} \quad (8)$$

where $\tilde{\mathbf{X}} = [\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_j, \dots, \tilde{x}_P]$ is the subset of P ($P \leq C$) horizontal coordinates such as $\bar{\mathbf{S}}(\tilde{\mathbf{X}}, y'_{\text{hip}}, t_i) = 1$.

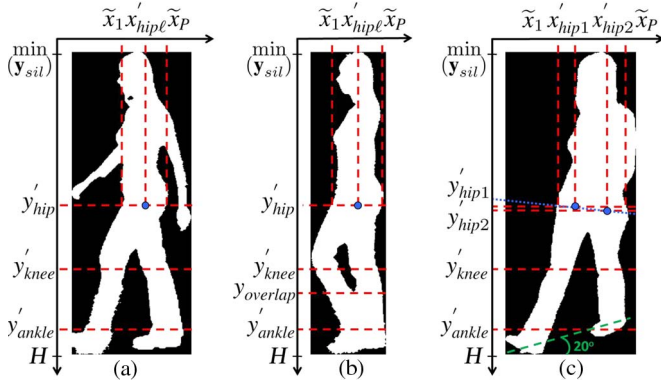


Fig. 1. Markerless gait analysis: $\mu = 0^\circ$ view with (a) nonoverlapped and (b) overlapped feet and (c) $\mu = 20^\circ$ view.

Equation (8) puts in relationship the horizontal hip position and the walking direction μ , calculated with respect to the horizontal axes of the image reference system. These relationships have been obtained with a regression analysis of the 3-D Georgia Tech motion capture data [48] by considering different camera positions.

The angle μ is estimated as the angle of inclination of the straight line which approximates the heel strike points with coordinates

$$(\mathbf{x}_{\text{str}}, \mathbf{y}_{\text{str}}) = (\mathbf{x}_{\text{feet}}, \mathbf{y}_{\text{feet}}) \left| \sum_{i=1}^P SK(\mathbf{x}_{\text{feet}}, \mathbf{y}_{\text{feet}}, t_i) > \tau \right. \quad (9)$$

where SK is the skeleton of the image, \mathbf{x}_{feet} and \mathbf{y}_{feet} are the coordinates corresponding to the portions of silhouette belonging to the feet, and the threshold τ is automatically set so that at least four heel strikes are extracted (in accordance with the hypothesis listed before).

Subsequently, a finer hip pose estimation is obtained with a linear approximation of the thighs by the first-order polynomial with coefficients

$$q_{\ell 0}(t_\varphi) = \frac{y_{\text{shin}^\ell}^m \cdot x'_{\text{hip}^\ell} - y'_{\text{hip}^\ell} \cdot x_{\text{shin}^\ell}^m}{x'_{\text{hip}^\ell} + x_{\text{shin}^\ell}^m}$$

$$q_{\ell 1}(t_\varphi) = \frac{1}{2} \cdot \left(a'_\ell + \frac{y'_{\text{hip}^\ell} - q_{\ell 0}}{x'_{\text{hip}^\ell}} \right) \quad (10)$$

where $y_{\text{shin}^\ell}^m = \min(\mathbf{y}_{\text{shin}})$, and $x_{\text{shin}^\ell}^m$ is the mean value of the correspondent horizontal coordinates. a'_1 and a'_2 are the slopes of the straight lines that approximate the left and right edges of the positions of silhouette belonging to the thighs, respectively.

Therefore, the angle between the thighs and the vertical axis at every frame t_φ is

$$\gamma_\ell(t_\varphi) = \pi - \arctan[q_{\ell 1}(t_\varphi)]. \quad (11)$$

Ultimately, the thighs are labeled as “left” or “right” in accordance with the corresponding shin labeling. Fig. 1 shows the shin and thigh extraction procedure. The angles α_1 , α_2 , γ_1 , and γ_2 are then corrected by projecting the limb trajectories on the lateral plane, as explained next.

B. Viewpoint Rectification

The method for viewpoint rectification is based on four main assumptions [35], [49].

- 1) The nature of human gait is cyclic.
- 2) Subjects walk along a straight line segment.
- 3) The distances between the bone joints are constant.
- 4) The articulated leg motion is approximately planar.

Considering a subject walking along a straight line, the multiple periods of linear gait motion appear analogous to a single period viewed from many cameras related by linear translation. Following this rationale, the positions of the points of interest, i.e., the leg joints, lie in an auto-epipolar configuration consistent with the imaged motion direction. The epipole is thus estimated by computing the intersection of the set of lines formed by linking the correspondent points of interest in each phase of the gait cycle. In order to find these correspondences, the gait periodicity is calculated by applying the stereopsis transformation that maps the epipole \mathbf{e} to the ideal point $[1, 0, 0]^T$ and then by computing the cost based on dot product between matching limb segment vectors.

Let \mathbf{j}_i^ℓ be the set of joint positions for each leg $\ell = \{1, 2\}$ at the i th frame in the image reference system. After estimating the periodicity of gait, assuming linear velocity between consecutive frames, the set of points of interest \mathbf{j}_i^ℓ is recomputed in order to lie on straight lines starting from the epipole. At first, the set of points and the epipole are mapped to the unit square and renormalized to the unit norm $\|\mathbf{e}_0\| = 1$, respectively. Subsequently, the optimal points are found by estimating the positions \mathbf{j}_i^ℓ that lie on the epipolar line and that satisfy the condition

$$\mathbf{j}_i^{\ell T} [\mathbf{e}_0] \times \mathbf{j}_i^\ell = 0. \quad (12)$$

Therefore, the back-projected rays, formed from a set of optimal points, intersect in a single world-space point: the epipole. The back projection of all sets of points generates the cluster of 3-D points for an assumed single period of reconstructed gait motion. The direct linear transform is then used in order to triangulate each world-space point \mathbf{J}^ℓ

$$\mathbf{j}_i^\ell \times \mathbf{p}_i \cdot \mathbf{J}^\ell = 0 \quad (13)$$

with the set of camera projection matrices

$$\mathbf{P}_i = [\mathbf{R}_e^T, -i\mathbf{e}_0] \quad (14)$$

where \mathbf{j}_i^ℓ is the image of the world-space point \mathbf{J}^ℓ in the i th period image, \mathbf{R}_e^T is the three-by-three rotation matrix that aligns the epipolar vector with the X (horizontal) axis, and i is an integer describing the periodicity of the subject's translation.

Considering the assumption that the articulated leg motion is approximately planar, the 3-D limb points can be reasonably fitted to two planes. Since the epipolar vector is aligned with the X axis, the ideal point $[1, 0, 0, 0]^T$ lies on each of the world-space planes. Therefore, the pencil of planes that intersect this ideal point has the form $p = [0, v_2, v_3, v_4]^T$. Consequently, the problem is reduced to finding two lines within the YZ -plane cross-sectional data.

After computing the mean $[y, z]^T$ of the point distribution, the translation \mathbf{H}_t that maps this point to the origin is applied. The two cross-sectional plane lines $l_1 = [v_2, v_3, v_4]^T$ and $l_2 = [v'_2, v'_3, v'_4]^T$ are then achieved by orthogonal regression and then aligned parallel with the Y (vertical) axis by applying a rotation \mathbf{H}_r . The intersection point of the two lines is then called u and is given by the cross product between the two lines.

Consequently, the pair of transformed lines is mapped to $\mathbf{l}'_\ell = \mathbf{H}_r \cdot \mathbf{l}_\ell$, and the rotation matrix \mathbf{H}_r and perspective transformation \mathbf{H}_α

$$\mathbf{H}_\alpha = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \alpha & 0 & 1 \end{pmatrix} \quad (15)$$

are applied to the point \mathbf{u}' in order to transform it to the ideal point $[1, 0, 0]^T$.

Since \mathbf{u}' lies on the Y axis and has the form $[y, 0, w]^T$, the transformation $\mathbf{H}_\alpha \mathbf{u}'$ gives $\alpha = -w/y$, and the corresponding line mapping $\mathbf{H}_\alpha \mathbf{l}'_\ell$ effectively makes the first component of the two normal lines zero. Since the lines are parallel, they are normalized

$$\mathbf{l}'_1 = [0, 1, -c_1]^T \quad \mathbf{l}'_2 = [0, 1, -c_2]^T \quad (16)$$

so that it is possible to find the point (c_1, c_2) of intersection with Z (depth) axis. A further similarity transform \mathbf{H}_s that translates the midpoint $(c_1, c_2)/2$ to the origin and scales in the Z direction to rectify the lines to the form $l = [0, 1, \pm 1]^T$ is then applied.

The translation by ± 1 mapping the selected set of points onto the $Z = 0$ plane is then computed with the matrix \mathbf{H}_b . The combined set of transformations thus forms the limb plane transformation $\mathbf{H} = \mathbf{H}_\beta \mathbf{H}_s \mathbf{H}_\alpha \mathbf{H}_r \mathbf{H}_t$.

Therefore, the projection transform mapping the back-projected points into the image can be decomposed as

$$\mathbf{j}_i^\ell = \mathbf{P}_i \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{H}^{-1} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{H} \end{pmatrix} \mathbf{J}^\ell \quad (17)$$

where

$$\mathbf{H} = \begin{pmatrix} 1 & m_2 & m_3 & m_4 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -\beta \\ 0 & -\alpha & 0 & 1 \end{pmatrix}. \quad (18)$$

The corresponding transformation of world-space points $[u, v, 0, w]^T$ into the image is given by

$$\mathbf{j}_i^\ell = \mathbf{H}_p [u, v, w]^T \quad (19)$$

where

$$\mathbf{H}_p = [\mathbf{e}' \quad m'_2 - \alpha \cdot (m'_4 - i \cdot \mathbf{e}') \quad (m'_4 - i \cdot \mathbf{e}') - \beta \cdot m'_3] \quad (20)$$

with

$$m'_i = \mathbf{R}_e^T \quad \mathbf{e}' = \mathbf{R}_e^T [1, 0, 0]^T. \quad (21)$$

Finally, the sets of optimal $Z = 0$ plane points are found by a solution to

$$\mathbf{x}'_{i,\beta} \times \mathbf{H}_p(i, \beta) \bar{\mathbf{w}} = 0 \quad (22)$$

for each point \mathbf{w} in order to minimize the reprojection error.

A structure on the $Z = 0$ plane has been recovered up to an affine ambiguity \mathbf{H}_μ that maps the imaged circular points $[1, \mu, \pm j\lambda, 0]^T$ back to their canonical positions $[1, \pm j, 0]^T$

$$\mathbf{H}_\mu = \begin{pmatrix} 1 & 0 & 0 \\ -\frac{\mu}{\lambda} & \frac{1}{\lambda} & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (23)$$

For estimating the metric structure, the lengths of the articulated limbs are assumed to be known and constant over all the frames. Thus, the squared distance between two points can be written

$$\mathbf{D}_\ell^2 = \Delta \mathbf{j}_i^{\ell T} \Delta \mathbf{j}_i^\ell \quad (24)$$

where

$$\Delta \mathbf{j}_i^\ell = [u_1 - u_0, v_1 - v_0]^T \quad (25)$$

if we consider the distance between points zero and one. If $\Delta \mathbf{j}_i^\ell$ and $\Delta \mathbf{j}_{i+1}^\ell$ are the pose difference vectors for a limb segment at two consecutive frames, then the equal limb length constraint can be written

$$\Delta \mathbf{j}_i^{\ell T} \cdot \mathbf{H}^T \cdot \mathbf{H} \cdot \Delta \mathbf{j}_i^\ell = \Delta \mathbf{j}_{i+1}^{\ell T} \cdot \mathbf{H}^T \cdot \mathbf{H} \cdot \Delta \mathbf{j}_{i+1}^\ell. \quad (26)$$

Therefore, writing $\Delta \mathbf{j}_i^{\ell T}$ and the element of the matrix $\mathbf{M} = \mathbf{H}^T \mathbf{H}$ as $m = [\mathbf{M}_{11}, \mathbf{M}_{12}, \mathbf{M}_{22}]$, the equation is

$$|\delta x_1^2 - \delta x_2^2 \quad 2(\delta x_1 \delta y_1^2 - \delta x_2 \delta y_2^2) \quad \delta y_1^2 - \delta y_2^2| m = 0. \quad (27)$$

Since m is defined up to scale, then a minimum of two corresponding pose constraints is required. All constraints formed from all sets of combinations of the same limb frame poses are stacked on each swing plane.

The rectification matrix \mathbf{H}_μ is formed from the extracted parameters of $\mathbf{H}^T \mathbf{H}$, where $\mu = -m_2/m_3$ and

$$\lambda = \sqrt{m_1/m_3 - \mu^2}. \quad (28)$$

The ideal epipole $[1, 0, 0]^T$ is then mapped by \mathbf{H}_μ to $[1, -\mu/\lambda, 0]^T$; thus, a rotation \mathbf{H}_r is necessary in order to align the epipole back along the X axis such that $\mathbf{H}_a = \mathbf{H}_r \mathbf{H}_\mu$ is the affine transform that recovers metric angles and length ratios on both planes. Points on the metric plane \mathbf{w} are then mapped into the image as

$$\mathbf{j}_i^\ell = \mathbf{H}_p \mathbf{H}_a^{-1} (\mathbf{H}_a \bar{\mathbf{u}}) = \bar{\mathbf{H}} \bar{\mathbf{w}}. \quad (29)$$

Scaling is then applied to both planes in order to transform each first limb segment to unit length. The mean set of limb lengths for both planes is estimated as \mathbf{d}, \mathbf{d}' . These lengths are related by the interplane scaling: $\mathbf{d}_i = \tau \mathbf{d}'_i$. A minimal solution to this set of linear equations requires at least one valid length correspondence within the set of limb segments. With \mathbf{H}_r now known, the optimal first limb segment length \mathbf{D}_1 on the first

plane can be evaluated. The scaling transform \mathbf{H}_s that maps \mathbf{D}_1 to the unit length and updates both sets of points and projection homographies is then calculated.

After recovering the frontoparallel structure of gait, the representation of the leg joint function $[\mathbf{J}_x^\ell(t), \mathbf{J}_y^\ell(t)]$ is found by fitting a modified Fourier series to the data with fixed fundamental frequency f_0 and period T

$$\mathbf{J}_x^\ell(t) = \nu_x t + \sum_{k=1}^n a_k \cos\left(2\pi k f_0 \left(t + \frac{(\ell - v_1)T}{2}\right) + \phi_k\right) + \mathbf{J}_{x0}^\ell \quad (30)$$

analogously for $\mathbf{J}_y^\ell(t)$. Thus, the projection of the leg joints on the lateral plane is obtained with an optimized procedure in the following way:

$$\check{\mathbf{J}}^\ell(t) = [h_1 \quad h_2 \quad h_3] g\left(t + \frac{(\ell - 1)T}{2} : f_0, \mathbf{D}_\ell, \nu_x, \nu_y, F\right) \quad (31)$$

where $g(t)$ is the bilateral Fourier series function with coefficients F and h as the values of the inverse normalization transform matrix.

Therefore, starting from a video sequence from a single camera and without any calibration, the proposed markerless system, in junction with [35] and [49], estimates the gait parameters projected on the lateral plane.

C. View-Invariant Gait Identification

The processing and derivation of ultimately discriminative gait features from this trajectory-based data are still an unsolved and challenging problem [22], [50], [51] due to the compound nature of gait motion inherent in the numerous variables associated with it, including kinematics, kinetics, and anthropometrics [52]. Various different ways of extracting discriminative features from gait sequences have been proposed for the purpose of people identification. An important issue in gait recognition is the derivation of appropriate features that can capture the discriminative individuality from the subject's gait. Such features should respond to crucial criteria, such as robustness and invariance to weather conditions, clothing, and operating conditions. They should also yield a good discriminability across different subjects.

In order to identify a subject by their gait, we derive the angular measurements, as well as the trunk spatial displacement, which best describe the gait kinematics. The use of angular motion is very common in gait analysis and recognition. The angles of the joints, including the hip and the knee, are considered the most important kinematics of the lower limbs. The angular and displacement measurements are taken from a single full gait cycle. The gait features can be taken from the averaging over different gait cycles, but because of the database limitation, only a single gait cycle is considered. Feature selection is employed to derive as many discriminative cues as possible while removing the redundant and irrelevant gait features which may degrade the recognition rate. It is practically infeasible to run an exhaustive search for all the possible combinations of features in order to obtain the optimal subset for recognition due

to the high dimensionality of the feature space. For this reason, we employed the adaptive sequential forward floating selection search algorithm [53]. The algorithm uses a validation-based evaluation criterion which is proposed to find the subset of features that minimize the classification errors, as well as ensure good separability between the different classes. In contrast to the voting scheme used in K -nearest neighbor (KNN), the evaluation function uses different weights w to signify the importance of the most nearest neighbors. The probability score for a sample s_c to belong to class c is expressed in the following:

$$f(s_c) = \frac{\sum_{i=1}^{N_c-1} z_i w_i}{\sum_{i=1}^{N_c-1} w_i} \quad (32)$$

where N_c is the number of instances in class c , and the weight w_i for the i th nearest instance is related to proximity as

$$w_i = (N_c - i)^2. \quad (33)$$

The value of z_i is defined as

$$z_i = \begin{cases} 1, & \text{if } \text{nearest}(s_c, i) \in c \\ 0, & \text{otherwise} \end{cases} \quad (34)$$

such that the $\text{nearest}(s_c, i)$ function returns the i th nearest instance to the sample s_c . The Euclidean distance metric is employed to find the nearest neighbors.

The CCR is computed using the KNN classifier with the leave-one-out cross-validation rule. The KNN rule is applied at the classification phase due to its low complexity and, hence, fast computation besides the ease of comparison to other methods. In the leave-one-out validation, every instance from the original sample is used for testing and is validated against the remaining observations. This is repeated for all the observations in the data set. The recognition rate is computed as the average of all validations.

IV. EXPERIMENTAL TESTS

The proposed method for view-independent gait biometrics has been tested on a large number of video sequences. Synthetic data have been specifically created for evaluating the accuracy of the system in extracting the gait features from the images. On the other hand, the capability of the proposed approach in identifying people's gait independently from the walking direction has been confirmed by applying the method to three publically available gait databases. Results on real data have been compared with the ones obtained by other approaches found in literature.

A. Synthetic Data

A quantitative validation of gait analysis has been obtained in a virtual environment: Synthetic video sequences (500×490 pixels; 30 ft/s) of a humanoid walking for one gait cycle have been created with Poser 7 (by e-frontier) in order to provide the ground truth for validation purposes. Different spatial resolutions, image qualities, and compressions have been considered.

TABLE I
ANGLE'S ERRORS (IN TERMS OF MEAN AND STANDARD DEVIATION) ON
SYNTHETIC GAIT WITH DIFFERENT SPATIAL RESOLUTIONS

| Image resolution/ mean sil's height | m_e (deg) | σ_e (deg) | Proc. time (s/frame) |
|--|-------------|------------------|-------------------------|
| 500x490/200 | 2.04 | 2.12 | 1.44 |
| 250x245/100 | 2.80 | 2.15 | 0.97 |
| 163x163/50 | 3.06 | 3.56 | 0.37 |

TABLE II
ANGLE'S ERRORS (IN TERMS OF MEAN AND STANDARD DEVIATION) ON
SYNTHETIC GAIT WITH DIFFERENT ZERO-MEAN GAUSSIAN NOISES

| σ_n / PSNR(dB) | m_e (deg) | σ_e (deg) |
|-----------------------|-------------|------------------|
| 0/ ∞ | 2.04 | 2.12 |
| 5/28.12 | 2.30 | 2.13 |
| 25/24.25 | 3.01 | 3.38 |
| 50/21.89 | 3.30 | 3.72 |
| 100/15.23 | 3.25 | 3.68 |
| 150/7.01 | 4.25 | 3.49 |

A range of spatial resolutions (250×245 and 163×163 pixels) have been considered by subsampling the original images with factors two and three. Table I shows the comparative results in terms of mean (m_e) and standard deviation (σ_e) of the joint's angles compared to the ground truth over the gait cycle.

The results, with a mean value of 2.6 ± 32.61 , are particularly encouraging and present the same magnitude to the ones obtained with 3-D markerless systems or 2-D complex-model-based ones [9], [12]. The image resolution and silhouette's height reduce the algorithm performance, as predicted. On the other hand, a 50-pixel-height silhouette still maintains an acceptable result. Furthermore, Table I shows the mean time for processing one frame (Matlab7 code processed with 2-GHz Intel CPU): The proposed approach allows one to extract the kinematic information with computational cost per silhouette's pixel lower than $2 \cdot 10^{-4}$ s.

In addition, to simulate background subtraction imprecision, zero-mean Gaussian noise has been added to the synthetic images. The standard deviation (σ_n) of the added noise varies from 0 (original data) to 150, corresponding to peak signal-to-noise ratio (PSNR) from ∞ to 7.01 dB.

The results in Table II show m_e and σ_e with respect to the different PSNRs. Obviously, the error increases with the level of noise added to the frames, but it is lower than 5° even at high noise levels. The noise robustness allows one to understand how the proposed method depends on an accurate silhouette extraction and, thus, to extend the approach in a real context.

Since images from a network of cameras can be compressed for data storage problems, different JPEG compressions have been compared. Seven quality factors from 100 to 5 have been used for compressing the synthetic images, allowing a compression ratio $CR = S_u/S_c$ between 40 to 146, where S_u and S_c are the sizes of the uncompressed and compressed images, respectively. Fig. 2 shows the m_e with respect to the image compression. The value obtained with uncompressed images is also reported. The mean error increases rapidly with the compression and then remains constant at about 3.3 pixels for JPEG qualities between 75 and 25.

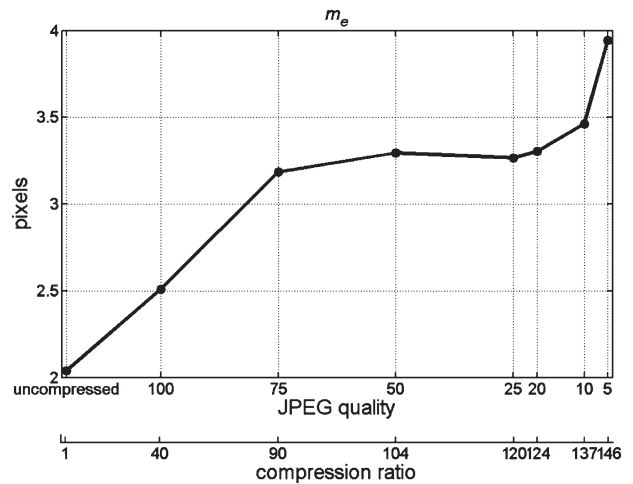


Fig. 2. Mean angle's errors on synthetic gait with different image compressions.

B. Real Data

The proposed method has been also tested on real data from two publically available databases, where a large number of subjects walk along different directions with respect to the camera.

The Southampton gait database has been used for comparing the proposed recognition approach with other methods found in literature to establish confidence in feature extraction. The CASIA-B database has been considered as the real test bed of the proposed view-invariant method since every single subject's walk is simultaneously captured from a large number of different views, and this establishes confidence in the viewpoint-invariant approach.

1) *Southampton Large Database*: For the evaluation of dynamic gait features derived using the proposed gait feature extraction method for people identification, a set of 160 video sequences is taken from the SOTON indoor gait database [19]. The set consists of 20 different subjects walking from left to right with eight sequences for every individual. There are ten males and ten females in the data set. Videos were recorded at a resolution of 720 by 576 pixels with people walking perpendicularly to the camera position. A sample of the markerless extraction of gait features is shown in Fig. 3.

From the extracted silhouettes, the joints' position has been estimated via the proposed markerless algorithm. Successively, the hip and knee angles have been used in the identification procedure. Fig. 3 shows an example of the feature extraction applied on a video sequence from the Southampton gait database.

The KNN rule is applied at the classification phase due to its low complexity and hence, fast computation besides the ease of comparison with some of the recent methods for gait biometrics. We have achieved a high recognition rate of 95.8% for the value of $k = 5$. This is achieved using solely features describing purely the dynamics of the locomotion process. The results of the recognition performance are summarized in Table III, with comparative results of other methods tested on similar subset of the SOTON database. From Table III, we can see that our method compares favorably with others. Moreover,

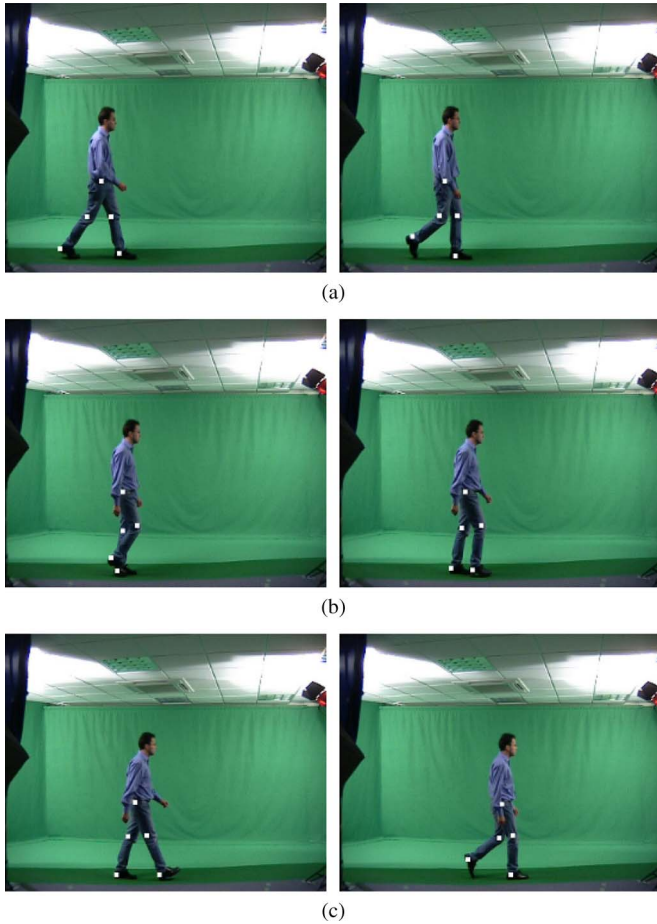


Fig. 3. Example images used in the experimental tests. (a) Small Southampton database. (b) CASIA-A database. (c) CASIA-B database.

TABLE III
COMPARISON OF CCRs (IN PERCENT) FROM SEVERAL DIFFERENT ALGORITHMS ON THE SOTON DATABASE

| Methods | CCR% |
|-------------------|-------------|
| Shutler 2000 [19] | 87.5 |
| Foster 2001 [21] | 83.0 |
| Wagg 2004 [54] | 78.0 |
| Yam 2004 [55] | 84.0 |
| Wang 2004 [56] | 87.5 |
| Chen 2007 [23] | 90.0 |
| Our method | 95.8 |

the recognition rate is above the ones found in literature and is particularly encouraging for tests on larger database, like CASIA-B. As such, we have great confidence in the accuracy of the markerless vertex extraction process.

2) *CASIA-B Database*: The view-invariant gait analysis has been tested on real data from a subset of CASIA-B database [24], where 124 young healthy subjects walk along straight lines with six different camera orientations (36° , 54° , 72° , 90° , 108° , and 126°). The 90° corresponds to the side-view walking direction, as shown in Fig. 4. Other extreme viewpoints are not considered in this experiment due to the fact that the motion of the legs can hardly be seen, and therefore, the detection of the gait features is not possible. The video sequences have spatial resolution and frame rate of 320×240 pixels and 25 ft/s, respectively.

The markerless feature extraction method has been applied to the CASIA-B gait data set consisting of 2270 video sequences for 65 different walking subjects with an average of six sessions for every viewpoint per subject. The limb pose has been estimated frame by frame, and the hip and knee angles have been extracted for each camera position and each subject. Fig. 4 shows an example of the limb pose estimation in the six different directions.

To investigate the effects of the viewpoint as a covariate factor for the performance of gait recognition, an initial experiment was carried out to measure the recognition rate using the nonrectified gait data. Successively, the viewpoint rectification method described in Section III-B is applied in order to reconstruct the gait angular features into the lateral plane (i.e., lateral view). Fig. 5(a) shows an example of the variations of hip angular motion during two gait cycles for the six different camera positions in the real experimental tests (CASIA B). Predictably, the angle trends, extracted in the image reference system, are influenced by the subject pose with respect to the camera, and they cannot be used directly for identification. For this reason, the viewpoint correction algorithm is applied, and the angle trends after the correction are shown in Fig. 5(b).

The CCR is computed separately for every viewpoint, and the experimental results are listed in Table IV. Results have been compared with the one obtained by Yu *et al.* with a silhouette-based approach [30].

In unrectified data, the CCR is first computed for all the data combined together, where a low CCR of 32% is reported based on the leave-one-out cross-validation. A set of extensive experiments is carried out to compute the recognition rates for every viewpoint separately. This is done based on probing various data sets of different and similar viewpoints. For the case of probing data sets of similar viewpoint to the gallery data set, high recognition rates are achieved, as shown in the diagonal of the table. The highest recognition rate is 87.5%, achieved when probing the side-view data set against the same viewpoint gallery data, while the recognition rate decreases when the walking direction changes from the lateral view. This is because of the variations in the extracted angular features from different views. Furthermore, the recognition rates drop significantly when matching data sets of different viewpoints. This is because the derived angular features are different. Therefore, nonrectified gait angular data cannot be used directly for biometric applications.

In the second experiment, the viewpoint rectification method described in the previous section is applied to reconstruct the gait angular features into the lateral plane (i.e., side view). Based on the same large data set of video sequences of walking subjects, a high recognition rate of 73.6% is achieved for all the viewpoints combined together for the rectified data, which compares well with the 32% achieved for the unrectified data. This is to signify the importance of the gait reconstruction into translating gait as a biometric method into real world surveillance and monitoring applications. The achieved results are compared with the baseline algorithm proposed by Yu *et al.* [30], where a silhouette-based approach is used to for a viewpoint-invariant gait recognition using the same gait data set. The

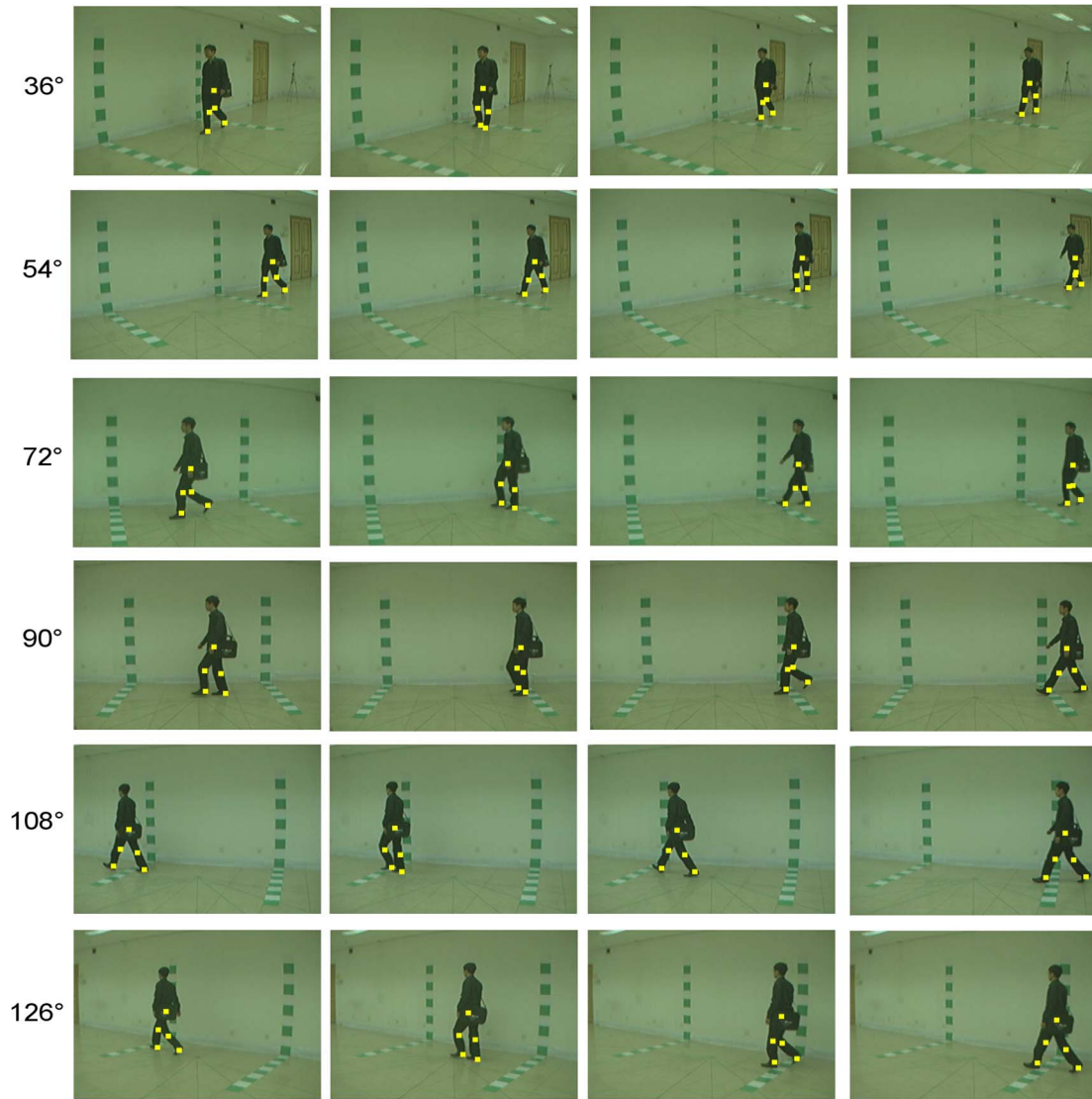


Fig. 4. Markerless joint extraction in different viewpoints.

results are shown in Table IV. A large decrease of the recognition performance is observed when probing two different viewpoints. This is mainly due to the nature of the silhouette-based approach, which is dependent on the subject shape that varies with the viewpoint.

Since the comparison between the two algorithms is not straightforward using these values in the three tables, two parameters (CCR_{nn} and CCR_{mn}) are designed for each experiment set. The first one is the distribution of recognition rates comparing the same views (values along the diagonal of Table IV). The second one is the distribution of recognition rates comparing different viewpoints (values apart from the diagonal). Comparative results, in terms of mean and standard deviation, are shown in Fig. 6.

Moreover, since recently, the CASIA-B database has been partially used for testing the method by Bashir *et al.* [29]; Fig. 7 shows the comparative results in terms of CCR (in percent) for probing different viewpoints against a gallery of data set recorded at the lateral view (90°). Both silhouette-based methods by Yu and Bashir are observed to have the

same pattern with a high recognition rate of over 90% when probing the lateral view against itself. However, a large drop of the recognition rate is reported when matching different viewpoints. For the case of the proposed joint-based gait recognition method, the CCR tends to be consistent over different viewpoints with an increased performance when matching the lateral viewpoint. This shows the benefit of using model-based approach combined with the rectification algorithm that can handle the effects of viewpoint. Clearly, the new approach allows for viewpoint-invariant analysis and handles practical factors in human movement analysis.

Another useful evaluation measure is the cumulative match score (CMS). The measure assesses the ranking capabilities of the recognition system by producing a list of scores that indicate the probabilities that the correct classification for a given test sample is within the top n matched class labels. Classification scores of 73.6% and 100% are achieved at the first and eleventh ranks, respectively. The CMS at the first rank is the CCR. Fig. 8 shows the CMS curve of the gait identification for the rectified and unrectified data.

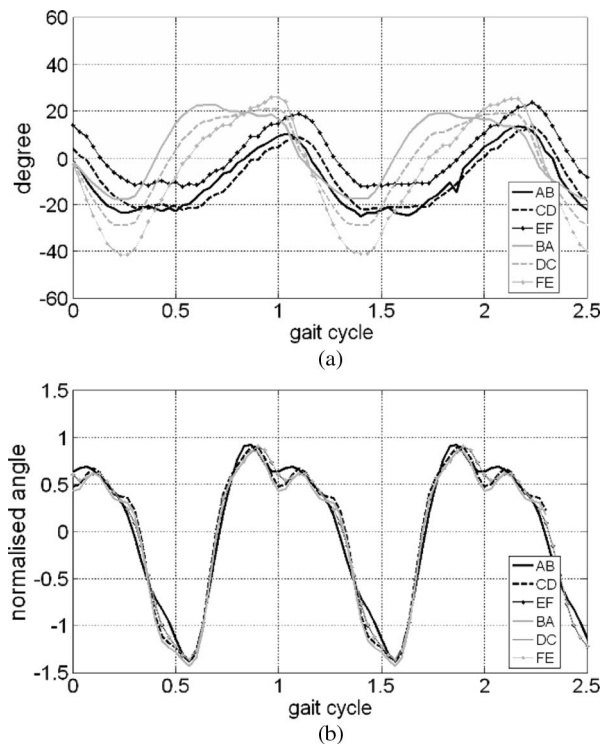


Fig. 5. Hip angular motion from different viewpoints. (a) Unrectified data. (b) Rectified angular data.

TABLE IV
CCR (IN PERCENT) CASIA-B, SET A. 1) UNRECTIFIED DATA.
2) RECTIFIED DATA. 3) RESULTS OF YU *et al.* [30]

| 1) Unrectified data | | | | | | | |
|--------------------------|------------------------------|------|------|------|------|------|------|
| Gallery Angle | Probe Angle (normal walking) | | | | | | |
| | | 36° | 54° | 72° | 90° | 108° | 126° |
| | 36° | 66.1 | 14.6 | 7.1 | 4.3 | 8.1 | 6.6 |
| | 54° | 13.7 | 72.9 | 17.6 | 10.3 | 8.1 | 8.3 |
| | 72° | 11.8 | 11.4 | 79.0 | 25.4 | 9.1 | 11.3 |
| | 90° | 8.1 | 10.8 | 28.6 | 87.5 | 21.3 | 11.9 |
| | 108° | 7.3 | 8.1 | 12.6 | 13.0 | 82.6 | 15.5 |
| | 126° | 6.9 | 8.7 | 8.8 | 9.2 | 20.4 | 77.4 |
| 2) Rectified data | | | | | | | |
| Gallery Angle | Probe Angle (normal walking) | | | | | | |
| | | 36° | 54° | 72° | 90° | 108° | 126° |
| | 36° | 72.1 | 56.5 | 53.1 | 53.0 | 46.7 | 44.6 |
| | 54° | 64.6 | 79.5 | 72.1 | 74.4 | 70.1 | 56.8 |
| | 72° | 59.7 | 65.1 | 85.0 | 72.6 | 59.1 | 60.1 |
| | 90° | 57.4 | 61.5 | 64.0 | 86.5 | 64.0 | 70.0 |
| | 108° | 57.3 | 63.4 | 69.7 | 69.2 | 82.3 | 67.6 |
| | 126° | 58.3 | 63.4 | 66.7 | 67.4 | 72.8 | 81.1 |
| 3) Yu <i>et al.</i> [30] | | | | | | | |
| Gallery Angle | Probe Angle (normal walking) | | | | | | |
| | | 36° | 54° | 72° | 90° | 108° | 126° |
| | 36° | 97.6 | 29.8 | 11.7 | 6.9 | 8.1 | 13.3 |
| | 54° | 29.0 | 97.2 | 23.0 | 16.5 | 21.4 | 29.0 |
| | 72° | 7.3 | 21.8 | 97.2 | 81.5 | 68.1 | 21.0 |
| | 90° | 4.8 | 17.7 | 82.3 | 97.6 | 82.3 | 15.3 |
| | 108° | 2.0 | 16.9 | 71.4 | 87.9 | 95.6 | 37.1 |
| | 126° | 6.0 | 37.5 | 33.5 | 22.2 | 48.0 | 96.8 |

V. CONCLUSION

We have taken an important step in deploying gait biometrics for the analysis of surveillance video. A view-invariant markerless model-based approach for gait biometrics has

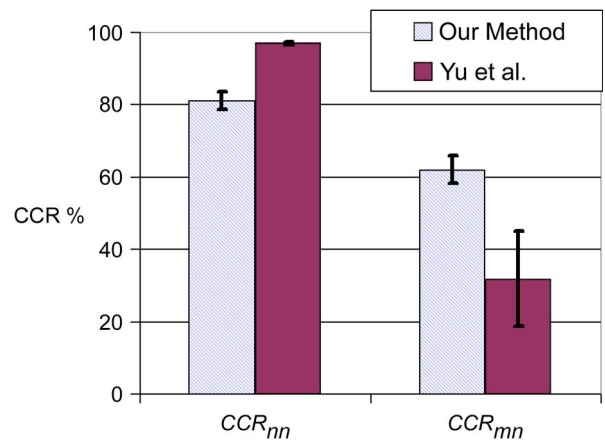


Fig. 6. CCR analysis for similar and different views. Comparison with the results from Yu *et al.* [30].

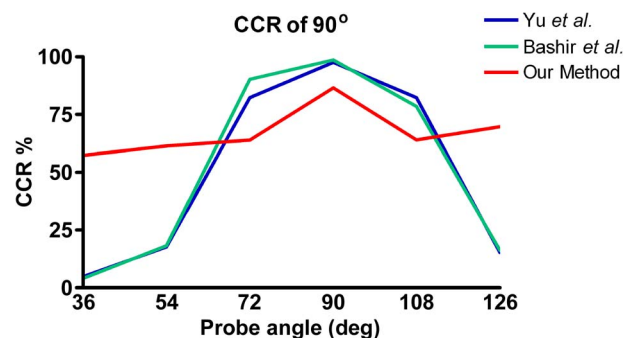


Fig. 7. Comparison of the gait recognition performance for probing different viewpoints against the lateral view.

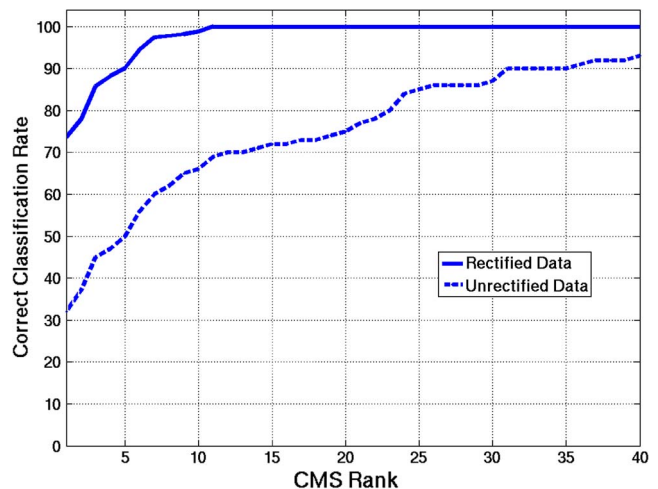


Fig. 8. CMS curve of the gait identification for the rectified and unrectified data.

been described. Gait features are derived based on the pose estimation of the joint positions of walking subjects. A novel 2-D markerless view-independent gait analysis algorithm has been presented: The method does not need camera calibration or prior knowledge of subject pose. Since the choice of the cameras' characteristics is a key point for the development of a smart surveillance system, the performance of the proposed approach has been measured with respect to different video properties.

A novel reconstruction method has been employed to rectify and normalize gait features recorded from different viewpoints into the side-view plane and therefore exploit such data for recognition. The method is used to investigate the performance of viewpoint-invariant gait recognition. A recognition rate of 73.6% is achieved using the KNN classifier with $k = 5$ being tested on a large gait data set with over 2000 video sequences. This concludes that people identification can be achieved by dynamic gait features, and we have encouraging results for our new view-invariant approach. This is an important step in translating gait biometrics into single-view scenarios where calibration information cannot be recovered such as in surveillance and forensic applications.

ACKNOWLEDGMENT

The authors would like to thank CBSR for providing access to the CASIA-B gait database.

REFERENCES

- [1] J. Cutting and L. Kozlowski, "Recognizing friends by their walk: Gait perception without familiarity cues," *Bull. Psychonom. Soc.*, vol. 9, pp. 353–356, 1977.
- [2] M. P. Murray, A. B. Drought, and R. C. Kory, "Walking patterns of normal men," *J. Bone Jt. Surg.*, vol. 46, no. 2, pp. 335–360, 1964.
- [3] M. S. Nixon and J. N. Carter, "Automatic recognition by gait," *Proc. IEEE*, vol. 94, no. 11, pp. 2013–2024, Nov. 2006.
- [4] I. Bouchrika and M. S. Nixon, *Model-Based Feature Extraction for Gait Analysis and Recognition*, vol. 4418, LNCS. New York: Springer-Verlag, 2007.
- [5] D. Cunado, M. S. Nixon, and J. N. Carter, "Automatic gait recognition via model-based evidence gathering," in *Proc. AutoID: IEEE Workshop Identification Adv. Technol.*, L. O'Gorman and S. Shellhammer, Eds., Oct. 1999, pp. 27–30.
- [6] G. Veres, L. Gordon, J. Carter, and M. Nixon, "What image information is important in silhouette-based gait recognition?" in *Proc. IEEE Conf. CVPR*, 2004, vol. 2, pp. II-776–II-782.
- [7] L. Wang, T. Tan, H. Ning, and W. Hu, "Silhouette analysis-based gait recognition for human identification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1505–1518, Dec. 2003.
- [8] P. Phillips, S. Sarkar, I. Robledo, P. Grother, and K. Bowyer, "The gait identification challenge problem: Data sets and baseline algorithm," in *Proc. Int. Conf. Pattern Recog.*, 2002, vol. 1, pp. 385–388.
- [9] R. Gross and J. Shi, "The CMU Motion of Body (MoBo) database," Robot. Inst., Carnegie Mellon Univ., Pittsburgh, PA, Tech. Rep. CMU-RI-TR-01-18, 2001.
- [10] J. Shutler, M. Grant, M. S. Nixon, and J. N. Carter, "On a large sequence-based human gait database," in *Proc. Int. Conf. Recent Adv. Soft Comput.*, 2002, pp. 66–72.
- [11] W. Hu, T. Tan, L. Wang, and S. Maybank, "A survey on visual surveillance of object motion and behaviors," *IEEE Trans. Syst., Man, Cybern.*, vol. 34, no. 3, pp. 334–352, Aug. 2004.
- [12] Y. Yamada, "Advanced method for improvement of obscure video image," in *Proc. IEEE 33rd Annu. Int. Carnahan Conf. Security Technol.*, 1999, pp. 440–445.
- [13] P. K. Larsen, E. B. Simonsen, and N. Lynnerup, "Gait analysis in forensic medicine," in *Proc. SPIE*, J.-A. Beraldin, F. Remondino, and M. R. Shortis, Eds., 2007, vol. 6491, p. 649 10M.
- [14] [Online]. Available: <http://news.bbc.co.uk/1/hi/england/lancashire/7343702.stm>
- [15] N. Lynnerup and J. Vedel, "Person identification by gait analysis and photogrammetry," *J. Forensic Sci.*, vol. 50, no. 1, pp. 112–118, Jan. 2005.
- [16] W. T. Dempster and G. R. L. Gaughan, "Properties of body segments based on size and weight," *Amer. J. Anatomy*, vol. 120, no. 1, pp. 33–54, Jan. 1965.
- [17] S. A. Niyogi and E. H. Adelson, "Analyzing and recognizing walking figures in XYT," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, Jun. 1994, pp. 469–474.
- [18] M. Constant and P. Ridgeon, *The Principles and Practice of CCTV*. Hertfordshire, U.K.: Paramount Publishing Ltd., 2000.
- [19] J. D. Shutler, M. S. Nixon, and C. J. Harris, "Statistical gait description via temporal moments," in *Proc. 4th IEEE Southwest Symp. Image Anal. Interpretation*, Mar. 2000, pp. 291–295.
- [20] J. B. Hayfron-Acquah, M. S. Nixon, and J. N. Carter, "Automatic gait recognition by symmetry analysis," in *Audio-and-Video-Based Biometric Person Authentication*, vol. 2091, LNCS, J. Bigun and F. Smeraldi, Eds. New York: Springer-Verlag, Jun. 2001, pp. 272–277.
- [21] J. P. Foster, M. S. Nixon, and A. Prudel-Bennett, "New area based metrics for automatic gait recognition," in *Proc. Brit. Mach. Vis. Conf.*, 2001, pp. 233–242.
- [22] L. Wang, T. Tan, W. Hu, and H. Ning, "Automatic gait recognition based on statistical shape analysis," *IEEE Trans. Image Process.*, vol. 12, no. 9, pp. 1120–1131, Sep. 2003.
- [23] S. Chen and Y. Gao, "An invariant appearance model for gait recognition," in *Proc. IEEE Conf. Multimedia Expo*, 2007, pp. 1375–1378.
- [24] [Online]. Available: www.cbsr.ia.ac.cn/english/Gait%20Databases.asp
- [25] J. Yang, X. Wu, and Z. Peng, "Gait recognition based on difference motion slice," in *Proc. 8th Int. Conf. Signal Process.*, Nov. 16–20, 2006, vol. 4, pp. 2428–2431.
- [26] S. Hong, H. Lee, I. Nizami, and E. Kim, "A new gait representation for human identification: Mass vector," in *Proc. 2nd IEEE ICIEA*, May 2007, pp. 669–673.
- [27] I. Nizami, S. Hong, H. Lee, S. Ahn, K.-A. Toh, and E. Kim, "Multi-view gait recognition fusion methodology," in *Proc. 3rd IEEE ICIEA*, Jun. 2008, pp. 2101–2105.
- [28] T. H. W. Lam, T. W. H. A. Ieong, and R. S. T. Lee, "Silhouette spatio-temporal spectrum (SSIS) for gait-based human recognition," in *Proc. Int. Conf. Adv. Pattern Recog.*, 2005, pp. 309–315.
- [29] K. Bashir, T. Xiang, and S. Gong, "Feature selection on gait energy image for human identification," in *Proc. IEEE ICASSP*, Mar. 31–Apr. 4, 2008, pp. 985–988.
- [30] S. Yu, D. Tan, and T. Tan, "A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition," in *Proc. 18th ICPR*, 2006, vol. 4, pp. 441–444.
- [31] C. BenAbdelkader, L. S. Davis, and R. Cutler, "Motion-based recognition of people in EigenGait space," in *Proc. IEEE Conf. FG*, 2002, pp. 267–274.
- [32] A. Y. Johnson and A. F. Bobick, "A multi-view method for gait recognition using static body parameters," in *Proc. Int. Conf. Audio-Video-Based Biometric Person Authentication*, 2001, pp. 301–311.
- [33] A. Kale, A. Chowdhury, and R. Chellappa, "Towards a view invariant gait recognition algorithm," in *Proc. IEEE Conf. AVSS*, 2003, pp. 143–150.
- [34] A. A. Kale, N. P. Cuntoor, B. Yegnanarayana, A. N. Rajagopalan, and R. Chellappa, "Gait analysis for human identification," in *Proc. AVBPA*, 2003, pp. 706–714.
- [35] M. Goffredo, N. Spencer, D. Pearce, J. N. Carter, and M. S. Nixon, *Human Perambulation as a Self Calibrating Biometric*, vol. 4778, LNCS. New York: Springer-Verlag, 2007.
- [36] R. Zhang, C. Vogler, and D. Metaxas, "Human gait recognition at sagittal plane," *J. Image Vis. Comput.*, vol. 25, no. 3, pp. 321–330, Mar. 2007.
- [37] J. E. Boyd, "Synchronization of oscillations for machine perception of gaits," *Comput. Vis. Image Underst.*, vol. 96, no. 1, pp. 35–59, Oct. 2004.
- [38] J. B. Hayfron-Acquah, M. S. Nixon, and J. N. Carter, "Automatic gait recognition by symmetry analysis," *Pattern Recognit. Lett.*, vol. 24, no. 13, pp. 2175–2183, Sep. 2003.
- [39] P. Huang, C. Harris, and M. Nixon, "Recognising humans by gait via parametric canonical space," *J. Artif. Intell. Eng.*, vol. 13, no. 4, pp. 359–366, Oct. 1999.
- [40] J. Shutler and M. S. Nixon, "Zernike velocity moments for sequence-based description of moving features," *Image Vis. Comput.*, vol. 24, no. 4, pp. 343–356, Apr. 2006.
- [41] D. K. Wagg and M. S. Nixon, "Automated markerless extraction of walking people using deformable contour models," in *Proc. Int. Conf. Comput. Animation Virtual Worlds*, 2004, vol. 15, pp. 399–406.
- [42] Z. Zhang and N. F. Troje, "View-independent person identification from human gait," *Neurocomputing*, vol. 69, no. 1–3, pp. 250–256, Dec. 2005.
- [43] A. Kale, A. Roychowdhury, and R. Chellappa, "Fusion of gait and face for human identification," in *Proc. IEEE ICASSP*, May 17–21, 2004, vol. 5, pp. V-901–V-904.
- [44] J. N. Carter and M. S. Nixon, "On measuring gait signatures which are invariant to their trajectory," *Meas. Control*, vol. 32, no. 9, pp. 265–269, Nov. 1999.
- [45] N. M. Spencer and J. N. Carter, "Viewpoint invariance in automatic gait recognition," in *Proc. 3rd IEEE AutoID*, 2002, pp. 1–6.
- [46] N. Spencer and J. Carter, "Towards pose invariant gait reconstruction," in *Proc. IEEE Conf. ICIP*, 2005, vol. 3, pp. III-261–III-264.

- [47] R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision*, vol. 1. Reading, MA: Addison-Wesley, 1992.
- [48] [Online]. Available: www.cc.gatech.edu/cpl/projects/hid
- [49] M. Goffredo, R. D. Seely, J. N. Carter, and M. S. Nixon, "Markerless view independent gait analysis with self-camera calibration," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recog.*, Sep. 2008, pp. 1–6.
- [50] B. Guo and M. Nixon, "Gait feature subset selection by mutual information," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 39, no. 1, pp. 36–46, Jan. 2009.
- [51] L. Lee and W. Grimson, "Gait analysis for recognition and classification," in *Proc. 5th IEEE Int. Conf. Autom. Face Gesture Recog.*, 2002, pp. 148–155.
- [52] T. Chau, "A review of analytical techniques for gait data. Part 1: Fuzzy, statistical and fractal methods," *Gait Posture*, vol. 13, no. 1, pp. 49–66, Feb. 2001.
- [53] P. Pudil, J. Novovičová, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognit. Lett.*, vol. 15, no. 11, pp. 1119–1125, Nov. 1994.
- [54] D. K. Wagg and M. S. Nixon, "On automated model-based extraction and analysis of gait," in *Proc. 6th IEEE Int. Conf. Autom. Face Gesture Recog.*, 2004, pp. 11–16.
- [55] C. Y. Yam, M. S. Nixon, and J. N. Carter, "Automated person recognition by walking and running via model-based approaches," *Pattern Recog.*, vol. 37, no. 5, pp. 1057–1072, May 2004.
- [56] L. Wang, H. Ning, T. Tan, and W. Hu, "Fusion of static and dynamic body biometrics for gait recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 2, pp. 149–158, Oct. 2004.



Michela Goffredo (M'07) was born in Rome, Italy, on May 12, 1978. She received the Electronic Engineering degree from the University of Rome "Tre," Rome, in 2003, with a thesis on markerless human posture analysis, and the Ph.D. degree in biomedical electronics from the Department of Applied Electronics, University of Rome "Tre," where she was involved in research on image processing for human movement analysis.

She is with the School of Electronics and Computer Science, University of Southampton, Southampton, U.K., working on gait biometrics, and also with the Department of Applied Electronics, University of Rome "Tre," working on arm motor control. Her research interests are image and video processing and human motion analysis.



Imed Bouchrika (M'07) received the B.Sc. and Ph.D. degrees in electronics and computer science from the University of Southampton, Southampton, U.K., in 2004 and 2008, respectively.

Since 2008, he has been a Research Fellow with the Information: Signals, Images, Systems Research Group, University of Southampton. His research areas are image processing, gait biometrics, and visual surveillance.



John N. Carter (M'90) received the B.A. degree in experimental physics from Trinity College, Dublin, Ireland, and the Ph.D. degree in astrophysics from the University of Southampton, Southampton, U.K.

In 1985, he changed discipline and joined the School of Electronics and Computer Science, University of Southampton, as a Lecturer researching in signal and image processing, where he is currently a Senior Lecturer with the Information: Signals, Images and Systems Research Group. In the past, he worked on programs as diverse as diesel engine diagnostics and vocal tract imaging. A recent success in this field is the development of a new dynamic form of magnetic resonance imaging, which makes it possible to reconstruct high time resolution multiplanar views of the human vocal tract while a subject is repeating a short phrase. His current major interest is in the general area of 4-D image processing, i.e., analyzing sequences of images to extract both 2-D and 3-D features, exploiting coherence over the whole sequence, i.e., imposing simple smoothness and continuity constraints. This has applications in object tracking and feature detection, where it is tolerant of high levels of noise and missing data. This has found application in biometrics, particularly in automatic gait analysis.



Mark S. Nixon (M'04–A'05) received the B.Sc. and Ph.D. degrees from the University of Reading, Reading, U.K., in 1979 and 1983, respectively.

He is currently a Professor in computer vision with the University of Southampton, Southampton, U.K. His research interests are image processing and computer vision. His team develops new techniques for static and moving shape extraction, which have found application in automatic face and automatic gait recognition and in medical image analysis. His team was early workers in face recognition, later came to pioneer gait recognition, and more recently joined the pioneers of ear biometrics. Among research contracts, he was the Principal Investigator with John Carter on the Defense Advanced Research Projects Agency-supported project "Automatic Gait Recognition for Human ID at a Distance." He has chaired or had major involvement in many conferences (British Machine Vision Conference, Conference on Audio- and Video-Based Biometric Person Authentication, IEEE Face and Gesture, International Conference on Pattern Recognition, International Conference on Biometrics, and IEEE International Conference on Biometrics: Theory, Applications and Systems) and given many invited talks. His vision textbook, cowritten with Alberto Aguado, *Feature Extraction and Image Processing* (Academic Press) reached 2nd Edition in 2008. With Tieniu Tan and Rama Chellappa, their book *Human ID Based on Gait*, which is part of the new Springer Series on Biometrics, was published in 2005.

Dr. Nixon is a Fellow of the Institution of Engineering and Technology and the International Association for Pattern Recognition.