Multimed Tools Appl (2016) 75:1201–1221
DOI 10.1007/s11042-014-2364-9

Towards automated visual surveillance using gait for identity recognition and tracking across multiple non-intersecting cameras

Imed Bouchrika · John N. Carter · Mark S. Nixon

Received: 3 March 2014 / Revised: 13 September 2014 / Accepted: 7 November 2014 /
Published online: 21 November 2014

Abstract Despite the fact that personal privacy has become a major concern, surveillance technology is now becoming ubiquitous in modern society. This is mainly due to the increasing number of crimes as well as the essential necessity to provide secure and safer environment. Recent research studies have confirmed now the possibility of recognizing people by the way they walk i.e. gait. The aim of this research study is to investigate the use of gait for people detection as well as identification across different cameras. We present a new approach for people tracking and identification between different non-intersecting un-calibrated stationary cameras based on gait analysis. A vision-based markerless extraction method is being deployed for the derivation of gait kinematics as well as anthropometric measurements in order to produce a gait signature. The novelty of our approach is motivated by the recent research in biometrics and forensic analysis using gait. The experimental results affirmed the robustness of our approach to successfully detect walking people as well as its potency to extract gait features for different camera viewpoints achieving an identity recognition rate of 73.6 % processed for 2270 video sequences. Furthermore, experimental results confirmed the potential of the proposed method for identity tracking in real surveillance systems to recognize walking individuals across different views with an average recognition rate of 92.5 % for cross-camera matching for two different non-overlapping views.

Keywords Gait analysis · Gait biometrics · Markerless extraction

I. Bouchrika (✉)
Department of Electrical Engineering, University of Souk-Ahras, Souk-Ahras, Algeria
e-mail: imed@imed.ws

J. N. Carter · M. S. Nixon
Department of Electronics and Computer Science, University of Southampton,
SO17 2BJ, Southampton, United Kingdom

1 Introduction

Although personal privacy has emerged recently as a major concern, surveillance technology is now becoming ubiquitous in such a modern society. This is mainly because of the increasing number of crimes as well as the essential need to provide better and safer environment. The possibility of tracking people over a network of cameras is of great interest in surveillance and security applications. This is because of the challenging nature for human operators to work simultaneously on different video screens in order to track people of interest as well as analyse their behaviours across different places. Thus, it has become an essential requirement for researchers from the computer vision community to investigate visual-based alternatives to automate the process for identity tracking over different views as well as analysis of human activities. Recently, various approaches were published in the literature to accomplish this goal based on deploying basic features such as shape or colour information. However, their practical deployment in real application is very limited due to the complex nature of such problem. Alternatively, we present a novel method for identity tracking of people across different cameras by the way of their walking pattern, their gait.

The use of gait for people identification in security and surveillance systems has recently attracted researchers from the computer vision community. The suitability of gait as biometrics emerges from the fact that gait pattern can be captured and perceived from a distance as well as its non-invasive and less-intrusive nature [5, 18, 33]. Early studies conducted by Murray [32] revealed that gait might be a useful biometrics for people identification. A total of 20 feature components including ankle rotation, spatial displacement and vertical tipping of the trunk have been identified to render uniquely the gait signature for every person. Furthermore, the studies carried out by Johansson [24] on the perception of human motion using Moving Light Displays revealed that based on the joints motion, an observer can recognise the different types of human activities. The observer can even make a judgment about the gender of the subject and further identify the person identity if they are already familiar with their gait pattern. This leads to the conclusion that the human gait pattern might be a potential biometric for surveillance applications.

As recent research studies have demonstrated that gait can be used as a potential biometrics for surveillance and forensic systems [20, 27, 33], we propose in this research a new approach for people re-identification across different stationary cameras with non-overlapping views through the use of their gait pattern. The approach is based on marker-less feature extraction method to derive the positions of the joints for a walking subject from uncalibrated single camera. Initially, people are detected through the extraction of their heel strikes on the ground, whereby the stride and cadence parameters can be estimated easily. In contrast to earlier methods, our approach does not require the subject to walk in the sagittal view, as gait features can be extracted using automated markerless methods from a number of different viewpoints [15] with the exception for the near frontal or rear view angles. Afterwards, the gait signature is constructed through the results of the extraction for the joints' positions of walking people [4] into the sagittal plane by a rectification process [40]. This is to transform gait angular data from a particular viewpoint to the normal i.e. sagittal plane. In an unconstrained environment where camera information is not available, people are tracked from different viewpoints by matching their gait signature against a database of existing signatures. This way, we can identify people being seen in one camera view from data already derived from a different camera view. Therefore, it is possible to recognize and track people in non-intersecting cameras via the use of gait.

The remainder of this paper is organised as follows. The next section summarises the previous approaches for people detection as well as identity recognition and tracking using gait across different cameras. The theoretical description of the proposed marker-less approach for deriving gait-based tracking signature is presented in Section 3. Section 4 introduces the experimental results and analysis applied.

2 Related work

There is a considerable research within the literature from various disciplines that affirms the concept of people recognition by the way they walk. This section surveys the recent state of the art studies related to people detection, identification as well as tracking across multiple Field of Views (FOV) through the use of their rhythmic gait pattern covering the different methodologies employed for feature extraction.

2.1 Pedestrian detection

The detection of pedestrians is of prime importance for most surveillance systems as it lays the basis for subsequent phases of an automated visual system, such as identity tracking or activity recognition. However, detecting and tracking people using a single camera is still a challenging problem due to shadows, occlusion, noise and natural background clutter. Furthermore, the flexible structure of the human body which encompasses a wide range of possible motion transformations along with the different covariate factors such as clothing, exacerbate difficulties for developing an automated visual-based surveillance system.

Existing surveillance systems for pedestrian detection can be categorized into several types [19] according to the number of cameras (single or multiple camera) or based on their functionalities (tracking single , multiple people, etc.). Lipton et al. [29] presented a real-time surveillance system to classify moving objects into either human or vehicle based on the “dispersedness” measure. In their work, people are assumed to have a dispersedness value smaller than the one for vehicles. However, shape metrics can vary depending on the image size as well as the distance from camera. The W^4 [19] surveillance system proposed by Davis et al. utilizes an appearance model in order to track people whereby single or group of subjects are distinguished using a projection histogram. Each subject within the group is located through the tracking of their head. Furthermore, Viola et al. [42] developed a real-time pedestrian detection system using AdaBoost. The system uses both image intensity as well as motion information for the classification of moving blobs.

Wang [43] surveyed two types of features being used for the detection of pedestrians within surveillance applications: shape-based and motion-based features. The first type relies on the shape of human silhouettes such as the dispersedenss value [29], aspect ratio of bounding box, or just using simple shape features. For the motion-based features, the periodicity of human motion is considered as a strong cue for walking people detection. Meanwhile, Cutler [11] proposed a real-time method for measuring the periodicity value for periodic motion based on self-similarity. Javed et al. [23] described a simple measurement procedure based repeated internal motion.

2.2 Gait-based identification for tracking

Much of the interest for human gait analysis has been originated from orthopaedics, physical therapy, and rehabilitation science for the diagnosis and treatment of patients with walking abnormalities. As the human gait pattern has recently emerged as a potential biometrics, gait analysis has now become a challenging computer vision problem. Various research studies have been published recently that were primarily aimed to develop a vision system capable of overcoming the difficulties encountered during the derivation and tracking of human motion features. A number of methods were surveyed in [25, 34].

BenAbdelkader et al. [3] described a pose-free method where the walking subject is detected and tracked. An image template which corresponds to the person's motion blob is extracted at each frame. Afterwards, a self-similarity plot from the derived sequence of templates can be computed. Experimental results were applied on outdoor data of 44 different subjects with four sequences for each person that were recorded on two different days. A Correct Classification Rate (CCR) of 77 % is reported. The method was tested on indoor data of seven subjects walking on a treadmill, recorded from 8 different viewpoints (from 0° to 120°) and on seven different days. A correct classification rate of 78 % was reported for near-front-parallel views whilst an average CCR of 65 % is obtained for all views.

The pose-based methods which construct or deduce the lateral view from data acquired from different arbitrary viewpoints, are the most recent approaches to 2D view-independent gait recognition. This is because the lateral view has proven better recognition capability with many approaches being published [22]. The vision research group at the University of Maryland proposed 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 viewpoint provided that the system knows the camera calibration parameters [8]. The method was tested on 12 subjects walking along straight lines recorded at different viewpoints. Considering the gallery dataset containing people walking at lateral view, the sequences where people walk at arbitrary views were chosen as the probe dataset whereby the Receiver Operating Characteristic (ROC) was computed for each view.

The image processing research group at the University of Southampton has focused attention on 2D view invariant gait biometrics since 1999 [7] where a trajectory-invariant gait signature was proposed. The method of Carter et al. rectifies the variations in gait data by knowing the walking trajectory of the human striking positions and modelling the thigh as a simple pendulum. The approach was then taken further by Spencer el al. [40] to propose a pose invariant biometric signature which did not require knowledge of the person's trajectory. Experimental results performed on synthesized data revealed that simple pose correction for geometric targets generalizes well for objects on the optical axis. Recently, the proposed method have been applied on subjects wearing reflective markers to aid the extraction whereby the gait signature was successfully reconstructed from six different views [40] with no prior knowledge of the camera calibration parameters.

2.3 Tracking & handoff between multiple cameras

The majority of research studies proposed for tracking people or moving objects over multiple cameras are limited in a way that they require knowledge about the camera calibration parameters as well as overlapping fields to maintain correspondence between different views. Camera calibration is an expensive task and the availability of calibration parameters in real life surveillance applications is proven challenging to obtain. Other research studies relax the requirement for camera calibration data but still need the overlapping fields to maintain correspondences using different types of basic features such as colour [36] or geometrical features. Cai et al. [6] described an approach to track walking subjects from sequences of synchronized and calibrated cameras. The correspondence across frames from different cameras is established via a set of feature points within a Bayesian probability framework. People are tracked using a single camera view until the system deduces that the active camera will no longer have a good view of the person. Basic features used for tracking include geometric properties such as the height of the person.

Stein et al. [41] proposed a novel method that does not require prior knowledge about the camera calibration. The camera calibration parameters are estimated by observing motion trajectories within the scene. Meanwhile, Javed [26] described a simple system for tracking people across multiple un-calibrated cameras. Their approach is able to infer the spatial relationships between the camera fields of views and use such information to establish the correspondence between different perspective views of the same subject.

3 Gait & visual surveillance

3.1 Walking people detection using point proximity

The rhythmic pattern of human gait is considered as the main feature to distinguish walking people from other moving objects. To derive a set of basic features for detecting walking people, frame differencing is applied to produce a motion map image based on the change detection for the inter-frame difference. The motion map M_t at frame t is derived based on a pixel-wise variation through of a window of consecutive images from the video stream. Afterwards, an accumulation procedure is being applied on the resulting motion maps through dividing the map into a grid with smaller bins of size 10×10 pixels and afterwards summing the values in each bin. A threshold is then applied to the accumulated image. Connected Component Analysis is finally performed to extract the larger blobs which correspond to moving objects within the scene.

When we walk, the foot for the striking leg is stabilised for half a gait cycle. As a result, if we stack the resulting motion images on top of each other as one image, a dense area of points is detected in the region where the foot strikes the ground. To locate these areas, we have proposed a measure for point proximity in an image to find where the crowded or dense region in a given image. The value of the proximity at a given point p is dependent on the number of points within the neighbourhood area R_p and their corresponding distances from p. For simplicity, R_p is considered to be a square area with the centre point p, and width $2r$ (r is set experimentally to 10 based on the image size). To compute the proximity image, we compute the neighbourhood proximity d_p for the region R_p corresponding to the point p, such that d_p is also a square region with the same width as R_p. The computation is carried out in an iterative process starting from the boundaries of R_p. It calculates the nearness value of points with respect to the centre p and then it iterates inside and accumulates the previous computed values as given in the following equation:

$$\begin{cases} d_p^r = \frac{N_r}{r} \\ d_p^i = d_p^{i+1} + \frac{N_i}{i} \end{cases} \tag{1}$$

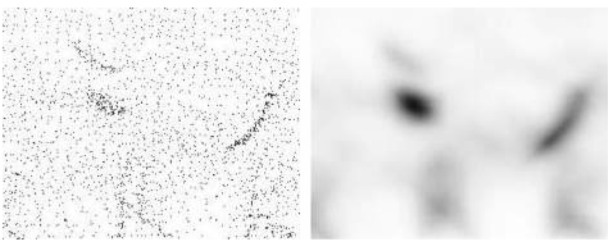
Such that d_p^i is the proximity value for squared rings of distance i away from the centre p, and N_i is the number of points at distance i from the centre.

Subsequently, to produce the final proximity image, we accumulate all the neighbourhood proximity values d_p for all points p into one image as outlined in the following equation:

$$D = \sum_{x=0}^{X}\sum_{y=0}^{Y} shift(d_{(x,y)}, x, y) \tag{2}$$

Such that X and Y are the width and height of the image frame respectively. $d_{(x,y)}$ is the neighbourhood proximity value for region $R_{(x,y)}$. The $shift$ function places the proximity value $d_{(x,y)}$ on an empty matrix of size $X \times Y$ at the position (x, y). An output of the point proximity for a given image is shown in Fig. 1. The input image contains a cloud of points placed at random with a number of dense regions. The resulting proximity image has darker areas which correspond to the crowded or dense regions within the input image. Because it is a difficult process to formally determine which regions in a given frame are dense as opposed to using a simple 2D histogram procedure, the problem becomes now a question of detecting darker regions of the derived proximity image. For this purpose, a similar algorithm to [14] is being employed to extract the positions of the peaks that correspond to crowded regions as local maxima.

For the context of surveillance within this study, the proximity measure is applied on different moving objects recorded using a surveillance camera. Moving objects include a single walking subject, a group of two people walking together and a vehicle as shown in Fig. 2. The proximity image derived from the motion maps of a walking person has a rhythmic pattern of darker spots being detected at the bottom part of the image as the foot hits the ground provided that the subject walks for at least two gait cycles. Furthermore, these dense areas are observed to have mostly the same level of darkness with consistent spatial distance between two consecutive regions. On the other hand, the proximity image for a group of people walking together constitutes a noisy pattern corresponding to the footsteps of subjects. Nevertheless, the bottom part of the proximity image is darker than the upper part of the proximity image. For moving vehicles, the proximity image has almost a flat pattern with peaks located at random positions. Furthermore, since gait is periodic, the stride length should be the same for different gait cycles, therefore the standard deviation of distances between two close strikes should tend to zero. For the classification of moving objects, we define a feature vector $< b, \sigma, \alpha >$ where b is the proportion of the lower part of the proximity image, and σ is the standard deviation value of distances between two successive peaks. α is the aspect ratio of height to width of the bounding box.



(a) (b)

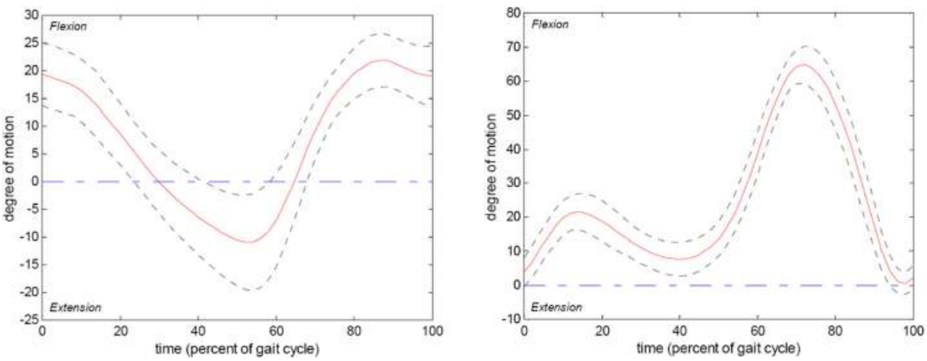
Fig. 1 The Point Proximity Measure: **a** Input Image, **b** Proximity Image

Fig. 2 People Detection using Gait Pattern

3.2 Markerless extraction of gait features

For recovering the lower limbs positions, a *Haar-like* feature template matching process is proposed for the localization of the legs using motion information for a single walking subject. The approach does not depend on background subtraction for the derivation of gait features. This is because it is computationally expensive and complex to deploy background subtraction for real-time surveillance applications due the process of updating the background model which is influenced by a number of factors such as background clutter, weather conditions and other outdoor environmental factors.

For the marker-less extraction of human gait feature, motion models are derived based on medical data that describes the angular motion for both the knee and hip at different phases of the gait cycle as shown in Fig. 3. The dashed lines represent the maximal and minimal points for the angular data of the human gait based on medical studies. A gait cycle is defined as the time interval between two successive instances of initial foot-to-ground contact of the same foot [10]. The hip initially bends by approximately 20° throughout the terminal stance phase and it extends afterwards until it reaches approximately 10° during the stance phase. During the pre-swing and throughout most of swing phase, the hip flexes



(a) (b)

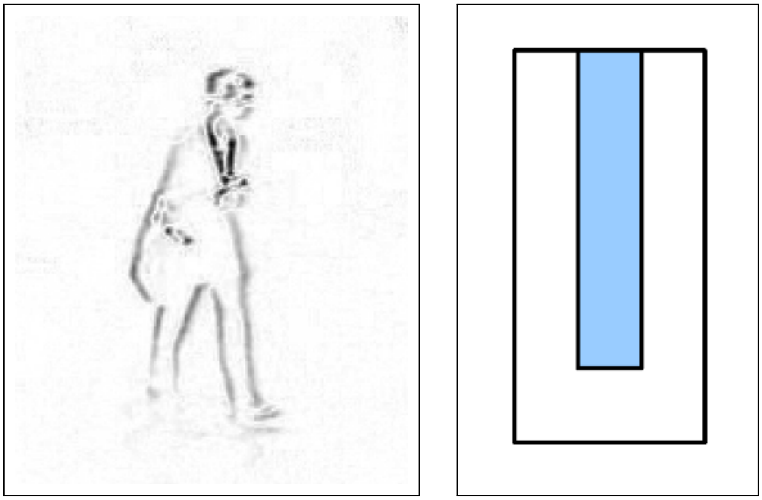
Fig. 3 Gait Angular Motion: **a** Hip. **b** Knee to nearly 20 degrees, and then it starts to extend just before the next initial contact with the ground. As shown in Fig. 3, the knee is almost fully extended during the first part of the mid-stance, it gradually begins to flex to its support phase peak which is about 20 degrees. The knee extends again almost fully and then flexes to approximately 40 degrees through the pre-swing phase. After toe-off, the knee flexes to reach a peak of 60 to 70 degrees at mid-swing, then it extends again in preparation for the next initial contact.

The proposed approach constructs initially the motion map image based on the change detection for the inter-frame difference. The only constraint of using frame differencing is that the camera must be in a stationary position. Moving pixels of a walking person across consecutive frames are detected with the emphasis to provide better edge data. The motion map M_t at frame t is estimated as the absolute difference of two consecutive frames I_t and I_{t+1} as given in the following equation:

$$M_t = \| I_t - I_{t+1} \| \tag{3}$$

Thresholding is applied on the resulting image M_t in order to reduce the artifacts. A sample motion image is illustrated in Fig. 4 for a walking subject recorded from a CCTV surveillance camera inside Gatwick airport.

A *Haar-like* template [35] is being utilized for the localization of the gait features due to their robust and fast performance in real-time systems from object recognition to pedestrian detection. The template is shown in Fig. 4 which is based on the outlier of the lower part of the leg. Let p_t^{ankle} is the candidate position of the ankle at t^{th} frame. To localize the ankle position, different templates are derived accounting for the different possible rotations and translation parameters defined by kinematical knowledge. The templates are superimposed against the motion map at the candidate point p estimating the match score S as given in (4). The position of the ankle is obtained once a maximum value for S is achieved. The similarity score describes how well is the matching template is superimposed on the motion map image. It is computed as the sum of larger values inside the superimposed region divided by



(a) (b)

Fig. 4 Markerless Gait Feature Extration: **a** Motion Image. **b** Haar-based Matching Template the accumulated lower values inside the area that are less than a certain threshold which is experimentally defined as $\tau = 20$.

$$S(x, y, \alpha) = \frac{\sum_{i \in P_{x,y,\alpha}} P_{x,y,\alpha}(i) \times Z(P_{x,y,\alpha}(i))}{\sum_{i \in P_{x,y,\alpha}} P_{x,y,\alpha}(i) \times |1 - Z(P_{x,y,\alpha}(i))|} \tag{4}$$

such that α is the rotation angle and Z is given as :

$$Z(i) = \begin{cases} 1 & \text{if } i > \tau \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

In contrast to using a per-frame algorithm for pose recovery [30], a frame-to-frame approach is being adopted for the extraction process where the results from the previous frame are exploited to guide the matching process in subsequent frames. Initially, the search process is performed over the whole motion region to find the best match for the leg. In order to limit the search space for a candidate point and refine further the extraction accuracy, kinematical and anthropometric constraints [15] including spatial as well as angular data derived from the gait motion models described in earlier section are imposed during the extraction process. For example, during the striking phase, one of the legs will be almost stabilised at the same position and therefore the ankle spatial movement is enclosed within a smaller region whilst the rotation parameter α will be limited within some specific range depending on the phase of the gait cycle. The lower limbs pose estimation algorithm is based on the proportions of the human body segments, following the medical results of anatomical studies [13] for a person of height H:

$$\begin{aligned} y'_{hip} &= \min\left(\mathbf{y}_{sil}\right) + 0.5 \cdot H \\ y'_{knee} &= \min\left(\mathbf{y}_{sil}\right) + 0.75 \cdot H \\ y'_{ankle} &= \min\left(\mathbf{y}_{sil}\right) + 0.90 \cdot H \end{aligned} \tag{6}$$

During the double-support phase of the gait cycle where the legs overlap, it is difficult to extract the lower limbs accurately because of the self-occlusion due to the overlap. Therefore, the matching process is applied for the striking leg using kinematic gait constraints that can assist with the localization. The swinging leg is skipped during the overlap. The overlapping phase starts whenever the Euclidean distance between the two ankle joints of both legs is less than a certain threshold which is related to the subject height. The extraction of the swinging leg during the overlap is resumed after a certain number of frames which is defined from the average gait cycle model. Experimentally, the number of frames is set to 6 for a video recorded with a frame rate of 25 frames/seconds. In order to extract the joints' positions as well the angular values α when the legs overlap, a 3^{rd} order polynomial interpolation has been applied. The choice of this polynomial has been determined experimentally.

The upper legs orientation is extracted for each frame $\mathbf{T} = \left[t_1, t_2, \ldots, t_\varphi, \ldots, t_F\right]$ with a coarse to fine hips estimation procedure where at first, the hips position is achieved with

$$\begin{cases} x'_{hip\ell} = \frac{1}{P} \cdot \sum\limits_{j=1}^{P} \widetilde{x}_j + (2\ell - 3) \cdot H \cdot \ \mu \cdot \ 10^{-3} \\ y'_{hip\ell} = y'_{hip} \cdot (2\ell - 3) \cdot \left(\frac{\widetilde{x}_P - \widetilde{x}_1}{2}\right) \cdot \sin(0.3 \cdot \mu) \end{cases} \tag{7}$$

such that $\widetilde{\mathbf{X}} = \left[\widetilde{x}_1, \widetilde{x}_2, \ldots, \widetilde{x}_j, \ldots, \widetilde{x}_P\right]$ is the subset P ($P \leq MaxX$) horizontal coordinates from extracted motion region for the subject S [16].

The (7) incorporates the relationship between 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 regression analysis of the 3D Georgia Tech motion capture data by considering different camera viewpoints. μ is estimated as the angle of inclination of the straight line which approximates the heelstrikes points.

3.3 View-point rectification

The approach proposed by Spencer et al. [40], is based on four main assumptions about the gait pattern in order to rectify the extracted data back into the normal plane: the nature of gait is cyclic; people usually walk in a straight line for a number of gait cycles; the distances between the human joints are constant; and the articulated leg motion is approximately planar. Thus, multiple periods of linear gait motion appear analogous to a single period viewed from multiple cameras related by linear translation and the positions of the joints lie in an auto-epipolar configuration.

If $\mathbf{j}_i^\ell$ is the set of positions for the joints for each leg $\ell = \{1, 2\}$ at the i^{th} frame in the image reference system, the relationship between $\mathbf{j}_i^\ell$ and the corresponding positions in the worldspace is $\mathbf{j}_i^\ell \times \mathbf{P}_i \cdot \mathbf{J}^\ell = 0$, where $\mathbf{P}_i = \left[\mathbf{R}_\mathbf{e}^T, -i\mathbf{e}_0\right]$ and $\mathbf{R}_\mathbf{e}^T$ is the rotation matrix for aligning the epipolar vector $\mathbf{e}_0$ with the horizontal axis X. Then,

$$\mathbf{j}_i^\ell = \mathbf{P}_i \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{H}_\mathbf{V}^{-1} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{H}_\mathbf{V} \end{pmatrix} = \mathbf{H} \cdot \mathbf{J}^\ell \tag{8}$$

The limb plane transformation matrix is expressed with $\mathbf{H}_\mathbf{V}$ so that the two cross section plane lines are normalised and centred with respect to Y and Z axes and parallel with Y. If we assume that the lengths of the articulated limbs $\mathbf{D}_\ell^2 = \Delta \mathbf{j}_i^{\ell\mathbf{T}} \Delta \mathbf{j}_i^\ell$ are constant across all the frames, the pose difference vectors for the limb segments at two consecutive frames, $\Delta \mathbf{j}_i^\ell$ and $\Delta \mathbf{j}_{i+1}^\ell$, are related as given below:

$$\Delta \mathbf{j}_i^{\ell\mathbf{T}} \cdot \mathbf{H}^\mathbf{T} \cdot \mathbf{H} \cdot \Delta \mathbf{j}_i^\ell = \Delta \mathbf{j}_{i+1}^{\ell\mathbf{T}} \cdot \mathbf{H}^\mathbf{T} \cdot \mathbf{H} \cdot \Delta \mathbf{j}_{i+1}^\ell \tag{9}$$

After reconstructing the fronto-parallel structure of the human gait, the representation of the leg joints function $\left[\mathbf{J}_x^\ell(t), \mathbf{J}_y^\ell(t)\right]$ is obtained by fitting a modified Fourier series to the data with fixed fundamental frequency f_0 and period T:

$$\mathbf{J}_x^\ell(t) = v_x t + \sum_{k=1}^{n} A_k \cos\left(2\pi k f_0 \left(t + \frac{(\ell-1)\,T}{2}\right) + \phi_k\right) + \mathbf{J}_{x0}^\ell \tag{10}$$

analogously for the value of $\mathbf{J}_y^\ell(t)$. Therefore, the projection of the leg joints on the sagittal plane is obtained with an optimized procedure as :

$$\breve{\mathbf{J}}^\ell(t) = \left[\, h_1 \;\; h_2 \;\; h_3 \,\right] g\left(t + \frac{(\ell-1)\,T}{2} : f_0, \mathbf{D}_\ell, v_x, v_y, F\right) \tag{11}$$

such that $g(t)$ is the bilateral Fourier series function with coefficients F and h are the values of the inverse normalization transform matrix [40]. Therefore, given a video stream from a single camera and without any prior knowledge about the calibration parameters, the proposed marker-less system, in conjunction with the work proposed by Spencer [40], we can estimate the gait parameters projected on the lateral plane derived from different viewpoints.

3.4 Derivation of gait signature

The gait biometric signature of a walking subject is composed from the magnitude and phase of the Fourier components from the extracted and rectified angular data taken for one full gait cycle. The phase information has a certain degree of importance in describing the dynamics of the human gait pattern. This is because the phase provides the information that describes when the gait dynamics occur. In order to compare phase vectors of different people, all analyses must be synchronized to start from the same point of the gait cycle. This point is chosen as the heel strike of the left leg. Since the magnitude data has been shown to offer low discriminatory capability even though it has the advantages of translation invariance property [9], the element-wise product of magnitude and phase is also used to construct the gait signature. Therefore, the gait signature is derived as the concatenation of magnitude and phases features along with the element-wise product of phase to magnitude features as given in (12).

$$f = (Magnitudes\ Phases\ Magnitudes \times \cdot\ Phases) \tag{12}$$

where $\times\cdot$ denotes the element-wise multiplication of magnitude and phase vectors, which weighs phase by magnitude to retain proportionate discriminatory ability. The total number of features in f is 675.

Feature selection is considered within this research studies to derive as many discriminative characteristics as possible whilst removing the redundant and irrelevant features which may degrade the recognition rate. It is practically impossible to run an exhaustive search for all the possible combinations of feature subsets in order to derive the optimal subset due to the high dimensionality of the features space. For this reason, the Adaptive Sequential Forward Floating Selection (ASFFS) search algorithm [39] is being applied.

The feature selection procedure is purely based on an evaluation procedure that assesses the usefulness or discriminativeness of each feature in order to obtain the best subset of features for the classification process. We have proposed validation-based evaluation criterion to select the subset of features that should minimise the classification errors as well as ensure good inter-class separation between the different clusters. As opposed to the voting scheme used by the k-nearest neighbors algorithm *KNN*, the evaluation procedure employs different coefficients w that signify the importance of most nearest neighbours of the same cluster. The probability score for a sample s_c to belong to a class c is described in the following (13):

$$f(s_c) = \frac{\sum_{i=1}^{N_c-1} z_i w_i}{\sum_{i=1}^{N_c-1} w_i} \tag{13}$$

such that N_c is the number of instances within class c, and the weight w_i for the i^{th} nearest instance is inversely related to proximity as given:

$$w_i = (N_c - i)^2 \tag{14}$$

The value of z_i is defined as:

$$z_i = \begin{cases} 1 & \text{if } nearest(s_c, i) \in c \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

where the *nearest*(s_c, i) function returns the i^{th} nearest instance to the sample s_c. The Euclidean distance metric is employed to determine the nearest neighbours from the same cluster. The significance for a subset of features is based on the validation-based metric which is estimated using the leave-one-out cross-validation rule.

4 Experimental results

4.1 People detection

In order to verify the effectiveness of the proposed method to classify moving objects using the gait pattern, experiments were carried out on the PETS2001 video dataset [12] containing a total of 27 moving objects being annotated from the set of four videos. The leave-one-out validation rule is being employed to examine the performance of the classification using the k-nearest neighbour (KNN) classifier. The algorithm is able to discriminate between a single walking person, a group of people and vehicles efficiently using the proposed motion features with an achieved Correct Detection Rate of 100 %. The feature vectors used for the classification of moving objects are projected into the feature space as shown in Fig. 5. Clearly the gait pattern is a strong cue to distinguish between a walking person from vehicles. This is consistent with the findings of BenAbdelkader et al. [2] where the gait stride parameter is used for people identification.

Further experiments were carried out on real surveillance video of the iLids 2009 dataset [21] provided by the UK Home Office. The dataset is acquired using CCTV cameras installed at Gatwick International Airport. For this experiment, the enhanced HoG method [31] for people detection was initially applied. Afterwards, the gait-based approach is applied to validate the existence of pedestrians through the use of gait motion. The main drawback of the proposed approach is its inability to perform well in crowded scenes due to the fact that gait features are occluded. When a subject is missed by the HoG detector provided that they are spotted in previous frames, the correspondence algorithm utilised for the temporal tracking of moving objects, is able to infer the missed detections. Table 1 shows a summary of people detection for the iLids dataset.

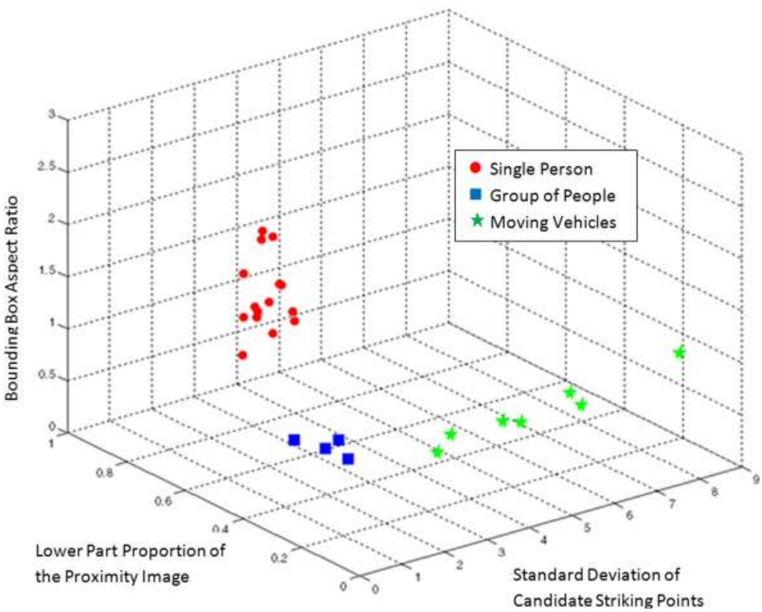


Fig. 5 Classification feature space

Table 1 Pedestrian Detection Results

	HoG Detection			Gait
	Missed	True	False	CCR
iLids2009	23	20	0	100 %

4.2 Gait recognition

In order to evaluate dynamic-based gait features derived using the marker-less extraction approach for gait recognition a gallery dataset of 160 video sequences is taken from the Southampton indoor gait database [38]. The set contains 20 different subjects with 8 videos for every person. The feature selection method is applied on the set of features derived from the image sequences in order to acquire the most discriminative subset of gait features. The correlation matrix is shown in Fig. 6 which visualizes the inter-class separation results across the different subjects based on the derived feature subset. The darker squares reflect higher separation score and thus higher discriminability. Meanwhile, the bright diagonal line reflects the zero distance between the same clusters.

To further evaluate the recognition potency of the selected feature subset, the Correct Classification Rate (CCR) is computed using the K-nearest neighbour classifier (KNN) with the leave-one-out cross-validation rule. The KNN is applied at the classification process due to its fast computation and simplicity besides the ease of comparison to existing methods. CCR of 95.7 % is obtained for $k = 5$ using the set of 160 video sequences. This is achieved using solely features describing purely kinematic-based features of the locomotion process derived from the angular data as expressed in (11). The results of the classification are being described in Table 2 with comparative results of other existing research studies which use dynamic-based features for gait recognition.

In order to assess the recognition performance of the proposed method using a different dataset, we took a different probe dataset from the Southampton gait database which was

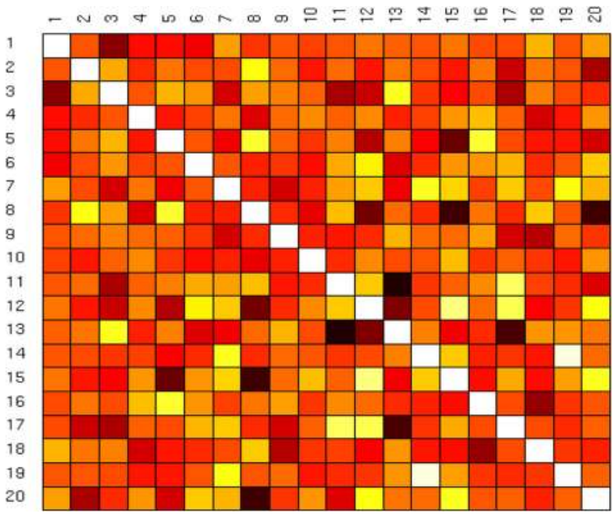


Fig. 6 Gait Recognition Correlation Matrix using the Southampton Dataset

Table 2 Classification Results of Gait Recognition

Method	Gait Database (SOTON)	CCR
Our method	20 subjects, 160 sequences	95.7 %
Yam [45]	20 subjects, 100 sequences	84 %
Wang [44]	20 subjects, 80 sequences	87.5 %

never used for the feature selection stage. The probe dataset is composed of 60 sequences for 20 candidates with 3 sequences for every person who exists in the gallery dataset. The probe dataset is thereafter matched against the gallery database. The Cumulative Match Score (CMS) evaluation approach [37] assesses the ranking capabilities of the recognition method by producing a list of score values that indicates the probabilities that the correct classification for a given test candidate is within the top n matched class labels. Using the CMS measure, we have obtained a correct classification rate of 86.67 % for the 60 walking sequences at rank $R = 1$. The achieved results using such evaluation are promising since the probe dataset has not been employed for the derivation of the dynamic feature subset. To conclude, dynamic features derived from angular data have a potential discriminative capability for people identification.

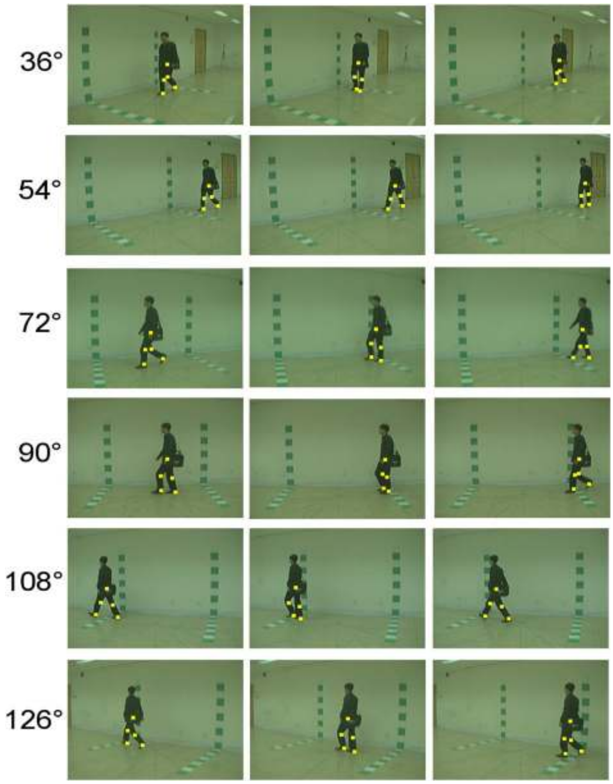


Fig. 7 Markerless Extraction of Gait Features

4.3 Viewpoint-invariant gait recognition

To assess the efficacy of the proposed approach for tracking walking people in multi-camera surveillance system, extensive experiments are performed for a variety of camera viewpoints. The Casia-B gait database [46] is considered as the real test-bed of the view-invariant algorithm. The walking pattern for every single subject is simultaneously captured from a large number of different views shown in Fig. 7. The marker-less feature extraction algorithm is run on batch mode to process the CASIA-B gait dataset consisting of 2270 videos for 65 different walking candidates with an average of 6 different sequences for every viewpoint per person. There are 6 different camera viewpoints which are: 36°, 54°, 72°, 90°, 108° and 126°. The 90° corresponds to the side view of the walking direction. The limbs pose is recovered on a frame by frame basis whilst the hip and knee angles are deduced for

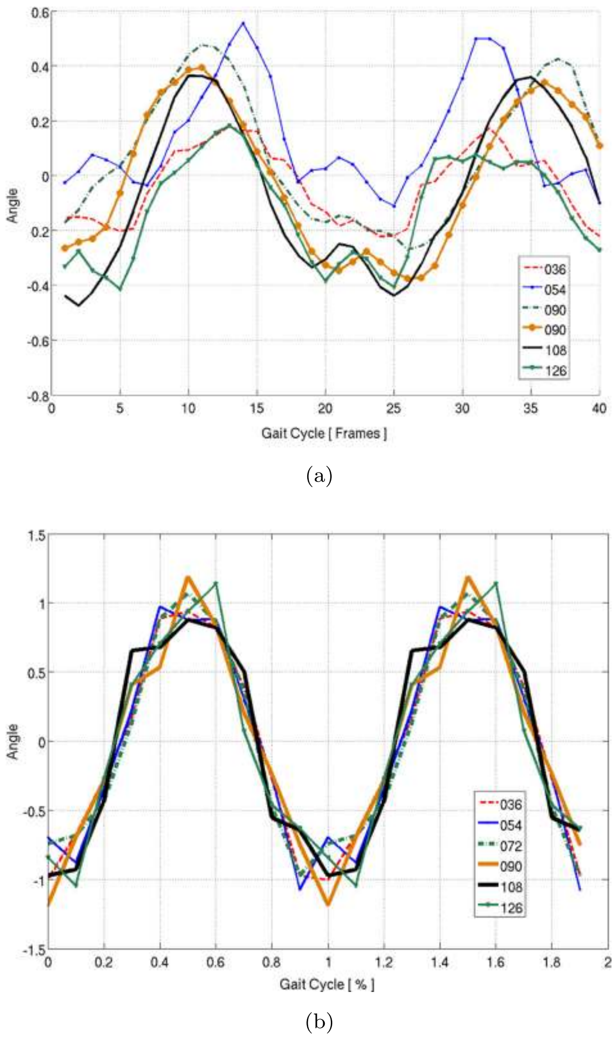


Fig. 8 Hip Angular Motion data from different View-points: **a** Unrectified. **b** Rectified

each camera position and for each person. Figure 7 shows an example of the extraction of the joints' positions for the 6 different viewpoints.

In order to investigate the effects of changing viewpoints as a covariate factor for the performance of gait biometrics, we have computed the correct classification rate using the selected subset of features derived from angular data but using non-rectified gait data. Subsequently, the viewpoint rectification process [40] described in Section 3.3 is applied to reconstruct the gait angular features into the normal plane (i.e. lateral view). Figure 8a shows an example of the variations for the hip angular data extracted during two gait cycles for the six different camera viewpoints from the CASIA-B dataset. Predictably, the angular data are influenced by the subject pose with respect to the camera position. Thus, they cannot be used directly for gait recognition. For this reason, the viewpoint rectification method is applied and the angular data for the human gait after the correction process are shown in Fig. 8b.

The CMS which was estimated for the Southampton dataset, is computed for the CASIA-B dataset for both the un-rectified and rectified angular data. For the rectified data, classification scores of 73.6 % and 100 % are obtained at the 1^{st} and 11^{th} ranks respectively as opposed to 32.0 % and 67.5 % for the un-rectified data. The CMS score at 1st rank is the correct classification rate. Figure 9 shows the CMS curve for the rectified and un-rectified data. In fact, the pose rectification process has contributed to increase the recognition process to reach 73.6 % for walking people recorded at different viewpoints which is the case of surveillance footage. The obtained results compare much better against the recognition rate reported by Yu et al. [46] for the same viewpoints of the CASIA-B achieving an average identification rate of 13.03 % using Gait Energy Image. Table 3 lists comparative results for gait recognition using mostly silhouette-based data being tested on the CASIA multi-view dataset. Although, the reported results for the silhouette-based approaches are better than the achieved results, the performance for silhouette-based methods can be highly affected under covariate factors such as clothing, footwear and load carriage [1, 4, 17, 46]

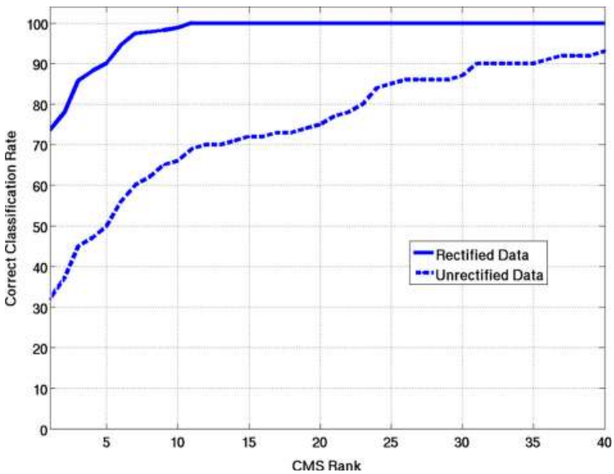


Fig. 9 CMS for Gait Recognition for Rectified and Un-Rectified data

Table 3 Classification Results of Gait Recognition under different Viewpoints

Method	Description	CCR
Our method	Without Rectification Process - KNN	32.0 %
Our method	Joints Motion / Rectification Process - KNN	73.6 %
Yu [46]	Gait Energy Image / KNN	13.0 %
Bashir [1]	Silhouette-base/Gaussian Process	86.5 %
Kusakunniran [28]	Silhouette Normalization/Procrustes distance	85.0 %

4.4 Identity tracking across multiple cameras

To determine the performance of the proposed approach using real life surveillance videos, we have evaluated the approach using the iLids dataset used earlier for people detection. The dataset is taken from CCTV surveillance cameras with overlapping and non-overlapping views. We have selected a set of 20 different walking people who are seen on different cameras views (Camera 2 and 3). The marker-less extraction algorithm is run to extract the angular data for the dataset of walking people in order to derive gait signatures from the gait and anthropometric features. Figure 10 previews an example of the extraction results of the joint positions for different camera views.

To investigate the efficiency of the proposed approach in such surveillance cases, we have used the leave-one-out cross validation with the KNN classifier to determine the performance across all the 20 subjects across two different non-overlapping cameras. The achieved correct classification rate (CCR) is 97 % for $k = 1$. Furthermore, data from

Fig. 10 Feature Extration applied on the iLids dataset

Table 4 Correct Classification Rate Analysis

	Camera 2 vs Camera 3	Camera 3 vs Camera 2
CCR	90 %	95 %

camera 3 are matched against data of camera 2 and vice versa in a probe to gallery mode. For the sake of imposing further challenges on the classification process, we have increased the size of the gallery dataset for camera 3 through adding 10 subjects that are recorded and processed within camera 3 only. In the same way, the number of subjects for the dataset of camera 2 is increased to include 10 more subjects. The obtained average correct classification is 92.5 % for the cross-camera matching process as shown in Table 4. This suggests that gait angular features can be used in surveillance systems for identity tracking and recognition across different cameras particularly for cases where it is impossible to derive robust features such as the face.

5 Conclusions

In this research study, we have taken an important step in deploying gait biometrics for the analysis of surveillance video. An approach for people tracking between different non-overlapping un-calibrated cameras based on gait analysis is being described. The identification signature is derived from gait angular data as well as anthropometric knowledge that are acquired using a marker-less feature extraction algorithm. Experimental results revealed the potency of our method to work in real surveillance systems to successfully recognize walking people over different views using the marker-less pose recovery with an average recognition rate of 92.5 % for cross camera matching. This is an important step in translating gait biometrics into real scenarios where prior knowledge about the camera calibration cannot be recovered such as in surveillance videos. It seems a natural avenue for future work to consider the complementarity of other gait biometrics approaches and to consider the benefits of other classification strategies. Furthermore, spontaneous gait detection and recognition of walking people in a crowded scene should be further investigated.

References

1. Bashir K, Xiang T, Gong S (2010) Cross view gait recognition using correlation strength. In: British Machine Vision Conference
2. BenAbdelkader C, Cutler R, Davis L (2002) Stride and Cadence as a Biometric in Automatic Person Identification and Verification. In: Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition, pp 357–362
3. BenAbdelkader C, Cutler R, Davis L (2002) View-invariant Estimation of Height and Stride for Gait Recognition. In: Proceedings of Workshop on Biometric Authentication
4. Bouchrika I (2008) Gait analysis and recognition for automated visual surveillance. University of Southampton. Ph.D. thesis
5. Bouchrika I, Carter J, Nixon M (2009) Recognizing People in Non-Intersecting Camera Views. In: Proceedings of Symposium on Imaging for Crime Detection and Prevention
6. Cai Q, Aggarwal J, Inc R, Seattle W (1999) Tracking human motion in structured environments using adistributed-camera system. IEEE Trans Pattern Anal Mach Intell 21(11):1241–1247

7. Carter JN, Nixon MS (1999) Measuring gait signatures which are invariant to their trajectory. Measurement Control 32(9):265–9
8. Chowdhury A, Kale A, Kale A (2003) Towards a view invariant gait recognition algorithm. In: Proceedings of IEEE Conference on Advanced Video and Signal Based Surveillance, pp 143–150. Citeseer
9. Cunado D, Nixon MS, Carter J (1997) Using gait as a Biometric, via Phase-Weighted Magnitude Spectra. In: Proceedings of 1st International Conference on Audio-and Video-Based Biometric Person Authentication, pp 95–102
10. Cunado D, Nixon MS, Carter JN (2003) Automatic Extraction and Description of Human Gait Models for Recognition Purposes. Comp Vision Image Underst 90(1):1–41
11. Cutler R, Davis LS (2003) Robust real-time periodic motion detection, analysis, and applications. IEEE Trans Pattern Anal Mach Intell 22(8):781–796
12. CVG (2001) Pets: Performance evaluation of tracking and surveillance. http://www.cvg.rdg.ac.uk/slides/pets.html
13. Dempster WT, Gaughran GR (1967) Properties of body segments based on size and weight. Am J Anat 120(1):33–54
14. Fujiyoshi H, Lipton AJ, Kanade T (2004) Real-time human motion analysis by image skeletonization. IEICE Trans Information System:113–120
15. Goffredo M, Bouchrika I, Carter J, Nixon M (2009) Self-Calibrating View-invariant Gait Biometrics. In: Journal of IEEE Transaction Systems, Man, and Cybernetics B
16. Goffredo M, Bouchrika I, Carter JN, Nixon MS (2010) Performance analysis for automated gait extraction and recognition in multi-camera surveillance. Multimedia Tools Appl 50(1):75–94
17. Guan Y, Li CT, Hu Y (2012) Robust clothing-invariant gait recognition. In: International Conference on Intelligent Information Hiding and Multimedia Signal Processing. IEEE, pp 321–324
18. Hadid A, Ghahramani M, Bustard J, Nixon M (2013) Improving gait biometrics under spoofing attacks. In: Image Analysis and Processing 2013. Springer, pp 1–10
19. Haritaoglu I, Harwood D, Davis LS (2000) W4: real-time surveillance of people and their activities. IEEE Trans Pattern Anal Mach Intell 22(8):809–830
20. Havasi L, Szlávik Z, Szirányi T (2007) Detection of gait characteristics for scene registration in video surveillance system. Image Processing. IEEE Transactions on 16(2):503–510
21. HOSD (2009) i-lids-imagery library for intelligent detection systems. https://www.gov.uk/government/collections/i-lids
22. Huang PS, Harris CJ, Nixon MS (1999) Human Gait Recognition in Canonical Space using Temporal Templates. In: IEE Proceedings of Vision, Image and Signal Processing, vol 146, pp 93–100
23. Javed O, Shah M (2002) Tracking and object classification for automated surveillance. In: Proceedings of the Seventh European Conference on Computer Vision, vol 4, pp 343–357
24. Johansson G (1973) Visual Perception of Biological Motion and a Model for its Analysis. Percept Psychophys 14:201–211
25. Katiyar R, Pathak VK, Arya K (2013) A study on existing gait biometrics approaches and challenges. Int J Comput Sci Issues 10(1)
26. Khan S, Javed O, Rasheed Z, Shah M (2001) Human tracking in multiple cameras. In: Eighth IEEE International Conference on Computer Vision, 2001. ICCV 2001. Proceedings, vol 1
27. Kim IS, Choi HS, Yi KM, Choi JY, Kong SG (2010) Intelligent visual surveillancea survey. Int J Control Autom Syst 8(5):926–939
28. Kusakunniran W, Wu Q, Zhang J, Ma Y, Li H (2013) A new view-invariant feature for cross-view gait recognition. Information Forensics and Security. IEEE Trans on 8(10):1642–1653
29. Lipton AJ, Fujiyoshi H, Patil RS (1998) Moving target classification and tracking from real-time video. 4th IEEE Workshop on Application of Computer Vision
30. Mori G, Malik J (2006) Recovering 3d human body configurations using shape contexts. IEEE Trans Pattern Anal Mach Intell 28(7):1052–1062
31. Morzinger R, Thaler M, Stalder S, Grabner H, Van Gool L (2011) Improved person detection in industrial environments using multiple self-calibrated cameras. In: Advanced Video and Signal-Based Surveillance (AVSS), 2011 8th IEEE International Conference on, pp 486–491. IEEE
32. Murray MP (1967) Gait as a Total Pattern of Movement. Am J Phys Med 46(1):290–333
33. Nixon MS, Bouchrika I, Arbab-Zavar B, Carter JN (2010) On use of biometrics in forensics: gait and ear. In: European Signal Processing Conference

34. Nixon MS, Tan TN, Chellappa R (2005) Human Identification Based on Gait. Springer, New York
35. Oren M, Papageorgiou C, Sinha P, Osuna E, Poggio T (1997) Pedestrian detection using wavelet templates. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp 193–199
36. Orwell J, Remagnino P, Jones G (1999) Multi-camera colour tracking. In: Second IEEE Workshop on Visual Surveillance, pp 14–21
37. Phillips PJ, Moon H, Rizvi SA, Rauss PJ (2000) The FERET Evaluation Methodology for Face Recognition Algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(10):1090–1104
38. Shutler J, Grant M, Nixon MS, Carter JN (2002) On a Large Sequence-Based Human Gait Database. In: Proceedings of Recent Advances in Soft Computing, pp 66–71
39. Somol P, Pudil P, Novovičová J, Paclík P (1999) Adaptive Floating Search Methods in Feature Selection. Pattern Recogn Lett 20(11–13):1157–1163
40. Spencer N, Carter JN (2005) Towards Pose Invariant Gait Reconstruction. In: Proceedings of IEEE International Conference on Image Processing, vol 3
41. Stein G (1999) Tracking from multiple view points: Self-calibration of space and time. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol 1, pp 521–527
42. Viola P, Jones MJ, Snow D (2005) Detecting pedestrians using patterns of motion and appearance. Int J Comput Vis 63(2):153–161
43. Wang L, Hu W, Tan T (2003) Recent Developments in Human Motion Analysis. Pattern Recog 36(3):585–601
44. Wang L, Ning H, Tan T, Hu W (2004) Fusion of Static and Dynamic Body Biometrics for Gait Recognition. IEEE Trans Circuits Syst Video Technol 14(2):149–158
45. Yam CY, Nixon MS, Carter JN (2004) Automated Person Recognition by Walking and Running via Model-Based Approaches. Pattern Recog 37(5):1057–1072
46. Yu S, Tan D, Tan T (2006) A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In: Proceedings of International Conference on Pattern Recognition, vol 4, pp 441–444

Imed Bouchrika received the Bsc and PhD degrees in Electronics and Computer Science from the University of Southampton (United Kingdom) in 2004 and 2008 respectively. Since 2008, he has worked as a research fellow at the Information: Signals, Images, Systems (ISIS) Research Group of the University of Southampton. He is now a lecturer of Computer Science at the University of Souk-Ahras. His research areas are image processing, gait biometrics and visual surveillance.

John N. Carter graduated with a BA in Experimental Physics from Trinity College, Dublin, Ireland followed by a PhD in Astrophysics from the University of Southampton England. In 1985 he changed discipline and joined the Department of Electronic Engineering also at Southampton as a lecturer researching in Signal and Image Processing. He is currently a Senior Lecturer in the Information: Signals, Images and Systems Research Group. In the past he has worked on programs as diverse as Diesel Engine Diagnostics and Vocal Tract imaging. A recent success in this field has been the development of a new dynamic form of Magnetic Resonance Imaging, which makes it possible to reconstruct high time resolution multi-planar views of the human vocal tract while a subject is repeating a short phrase. His current major interests are in the general area of Four-Dimensional Image Processing. That is, analysing sequences of images to extract both two and threedimensional 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, in particular Automatic Gait Analysis. Dr. Carter is a member of the IEEE.

Mark S. Nixon is the Professor in Computer Vision at the University of Southampton UK. His research interests are in 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 were early workers in face recognition, later came to pioneer gait recognition and more recently joined the pioneers of ear biometrics. Amongst research contracts, he was Principal Investigator with John Carter on the DARPA supported project Automatic Gait Recognition for Human ID at a Distance. He chaired BMVC 98 and with Josef Kittler he chaired the Audio Visual Biometric Person Authentication (AVBPA 2003), was Publications Chair for the International Conference on Pattern Recognition (ICPR 2004) and co-chaired the IEEE 7th International Conference on Face and Gesture Recognition FG2006. He has recently been program chair on BTAS'08 and ICB'09. His vision book, co-written with Alberto Aguado, Feature Extraction and Image Processing reached its Second Edition in 2007 and with Tieniu Tan and Rama Chellappa, his book Human ID based on Gait which is part of the new Springer Series on Biometrics, was published in 2005. Dr. Nixon is a member of the IEEE