

Covariate Analysis for View-point Independent Gait Recognition

I. Bouchrika, M. Goffredo, J. N. Carter and M. S. Nixon

ISIS, Department of Electronics and Computer Science
University of Southampton, SO17 1BJ, UK
`{ib04r,mg2,jnc,msn}@ecs.soton.ac.uk`

Abstract. Many studies have shown that gait can be deployed as a biometric. Few of these have addressed the effects of view-point and covariate factors on the recognition process. We describe the first analysis which combines view-point invariance for gait recognition which is based on a model-based pose estimation approach from a single un-calibrated camera. A set of experiments are carried out to explore how such factors including clothing, carrying conditions and view-point can affect the identification process using gait. Based on a covariate-based probe dataset of over 270 samples, a recognition rate of 73.4% is achieved using the *KNN* classifier. This confirms that people identification using dynamic gait features is still perceivable with better recognition rate even under the different covariate factors. As such, this is an important step in translating research from the laboratory to a surveillance environment.

1 Introduction

In recent years, automated visual surveillance has received considerable interest in the computer vision community. This is largely due to the vital need to provide a safer environment. Because of the rapid growth of security cameras and the need for automated analysis, the deployment of biometric technologies becomes important for the development of automated visual surveillance systems. The suitability of gait recognition for surveillance systems emerges from the fact that gait can be perceived from a distance as well as its non-invasive nature. Although gait recognition is not sufficiently mature to be used in real world applications such as visual surveillance, it overcomes most of the limitations that other biometrics suffer from such as face, fingerprints and iris recognition which can be obscured in most situations where serious crimes are involved.

Gait can be affected by different covariate factors including footwear, clothing, injuries, age, walking speed, and much more akin with other biometrics. In fact, the effects of the different covariates for gait analysis and recognition have not been investigated much by medical and other researchers [1]. This is mainly due to the lack of availability for databases, as well as the availability of automated systems which would help for the extraction of gait features. Moreover, the complexity of earlier model-based approaches has precluded their deployment for this analysis. The covariate factors can be related either to the

subject as for the case when a subject smiles for face recognition, or related to the environmental conditions such as lighting, nature of the ground or camera setup.

Much research for gait recognition has been done into identifying subjects recorded walking from the side-view. The effects of covariate factors on the performance of gait recognition have been investigated by only a few recent research studies. Sarkar *et al.* [2] described a baseline algorithm for gait recognition based on the temporal correlation of silhouette data. The algorithm is evaluated on a set of twelve experiments in order to examine the effects of the different covariates including viewpoint, footwear, walking surface, time and carrying conditions. However, their work lacks exploratory analysis of the different gait features under covariate data due to the use of the silhouette approach. Tan *et al.* presented an averaging silhouetted-based approach that was tested on the CASIA-B gait dataset with three main variations including clothing, carrying conditions and view angles [3]. Their experimental results showed that the performance of gait recognition is much affected at worst dropping to a recognition rate of just 1% for covariate dataset.

In this paper, a markerless model-based approach is used to investigate of the effects of the covariate factors including, clothing and carrying conditions for view-point independent gait recognition. This extends recent research studies by on covariate analysis [4] and view-point invariant gait recognition [5]. A novel reconstruction method is being employed to rectify and normalize gait features recorded from different view-point into the side-view plane and therefore exploit such data for recognition. As such, we show for the first time that covariate analysis and viewpoint invariance can be combined, thus handling important practical factors in the translation of gait from laboratory to surveillance analysis. This paper is structured as follows: the next section is devoted to the discussion of markerless method used for extracting gait features as well as the normalization approach used for reconstructing gait angular data into the side-view plane. Section 3 describes the gait recognition approach including the derivation of gait signatures and the classification process. Finally, the experimental results on a set of processed videos from CASIA dataset are drawn in the fourth section.

2 Automated Markerless Extraction of Gait Features

2.1 Estimation of the Joint Positions

To extract the gait features of walking subjects from the covariate dataset, we applied the model-based method described in [6] to automate the extraction process of the joint trajectories. To extract candidate joint positions, the Distance Transform is performed on the silhouettes of walking subjects. Spatial motion templates describing the motion of the joints are derived by manual gait analysis and used to aid the markerless extraction of the joint positions. A recursive evidence gathering algorithm is employed for the markerless extraction process whereby spatial model templates for the human motion are presented in

a parameterized form invariant to scaling and rotation using the Elliptic Fourier Descriptors described in equation (1):

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} a_0 \\ b_0 \end{bmatrix} + \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix} \begin{bmatrix} F_x(t) \cdot s_x \\ F_y(t) \cdot s_y \end{bmatrix} \quad (1)$$

where $t \in [0, 2\pi]$, α is the rotation angle, s_x and s_y are the scaling factors across the horizontal and vertical axes respectively. a_0 and b_0 define the position of the shape's centre. $F_x(t)$ and $F_y(t)$ are computed using equation :

$$\begin{aligned} F_x(t) &= \sum_{k=1}^n a_{x_k} \cos(kt) + b_{x_k} \sin(kt) \\ F_y(t) &= \sum_{k=1}^n a_{y_k} \cos(kt) + b_{y_k} \sin(kt) \end{aligned} \quad (2)$$

where $a_{x_k}, a_{y_k}, b_{x_k}$ and b_{y_k} are the set of the elliptic phasors which can be computed by Riemann summation [7]. Gait knowledge is exploited via heel strike extraction to reduce the the parameter space dimensionality and therefore reduce the computational load of the evidence gathering algorithm. The Hough Transform is employed to determine the free parameters through the matching process of feature points across the whole sequence of frames to the parametric function, and increase votes in the accumulator space accordingly. The parameters are then determined as the index or key of the accumulator space with the largest value. In the latter phase of the evidence gathering process, an exhaustive local search is performed within every frame to locate the features (i.e., joint positions) whereby, the local search is guided by the motion pattern extracted during the first stage to limit the search area. To more accurately extract the joint positions and reduce the search space, the lower limbs pose estimation algorithm uses as a filtering process the proportions of the human body segments.

2.2 View-Point Rectification

The rectification method is applied to normalise gait features extracted from any viewpoint into the side-view plane. The method is based on four main assumptions: the nature of human gait is cyclic; subjects walk along a straight line during two gait cycles; the distances between the bone joints are constant; and the articulated leg motion is approximately planar.

Considering a subject walking along a straight line, 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}_0 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 joints 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 are 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 re-normalized 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 satisfies the condition

$$\mathbf{j}_i^{\ell T} [\mathbf{e}_0] \times \mathbf{j}_i^\ell = 0 \quad (3)$$

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

$$\mathbf{j}_i^\ell \times \mathbf{P}_i \cdot \mathbf{J}^\ell = 0 \quad (4)$$

with the set of camera projection matrices

$$\mathbf{P}_i = [\mathbf{R}_e^T, -i\mathbf{e}_0] \quad (5)$$

\mathbf{R}_e^T is the rotation matrix for aligning the epipolar vector \mathbf{e}_0 with the horizontal axis X. Then,

$$\mathbf{j}_i^\ell = \mathbf{P}_i \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{H}_V^{-1} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{H}_V \end{pmatrix} = \mathbf{H} \cdot \mathbf{J}^\ell \quad (6)$$

having expressed the limb plane transformation matrix with \mathbf{H}_V so that the two cross section plane lines are centred and normalised respect to Y and Z axes and parallel with Y. By assuming the lengths of the articulated limbs $\mathbf{D}_\ell^2 = \Delta \mathbf{j}_i^{\ell T} \Delta \mathbf{j}_i^\ell$ are constant over 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 by

$$\Delta \mathbf{j}_i^{\ell T} \cdot \mathbf{H}^T \cdot \mathbf{H} \cdot \Delta \mathbf{j}_i^\ell = \Delta \mathbf{j}_{i+1}^{\ell T} \cdot \mathbf{H}^T \cdot \mathbf{H} \cdot \Delta \mathbf{j}_{i+1}^\ell \quad (7)$$

After recovering the fronto-parallel structure of subject gait, the representation of the leg joints 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 :

$$\mathbf{J}_x^\ell(t) = v_x t + \sum_{k=1}^n A_k \cos \left(2\pi k f_0 \left(t + \frac{(\ell-1) T}{2} \right) + \phi_k \right) + \mathbf{J}_{x0}^\ell \quad (8)$$

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

$$\mathbf{\dot{J}}^\ell(t) = [h_1 \ h_2 \ h_3] g \left(t + \frac{(\ell-1) T}{2} : f_0, \mathbf{D}_\ell, v_x, v_y, F \right) \quad (9)$$

where $g(t)$ is the bilateral Fourier series function with coefficients F and h are 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 estimates the gait parameters projected on the lateral plane.

3 Gait Recognition

The processing and derivation of good gait features from this trajectory-based data is a challenging problem due to the compound nature of gait motion inherent in the numerous variables associated with it including kinematics, kinetics and anthropometrics [8]. An important issue in gait recognition is the derivation of appropriate features that can capture the discriminative individuality from a subject's gait. Such features should respond to crucial criteria such as robustness and invariance to weather conditions, clothing and operating conditions.

In order to identify a subject by their gait, we derive the angular measurements, anthropometric measurements as well as the trunk spatial displacement which best describe the gait kinematics. The use of angular motion is very common in model-based 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 [9]. The anthropometric measurements include the subject height and lengths of the lower limbs. Feature selection is employed to derive as many discriminative cues as possible whilst 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 (ASFFS) search algorithm [10]. The algorithm uses a validation-based evaluation criterion which is proposed to find the subset of features that minimises the classification errors as well as ensure good separability between the different classes. In contrast to the voting scheme used in the KNN, the evaluation function uses different weights w to signify the importance of the most nearest neighbours. The probability score for a sample s_c to belong to class c is expressed in the following equation (10):

$$f(s_c) = \frac{\sum_{i=1}^{N_c-1} z_i w_i}{\sum_{i=1}^{N_c-1} w_i} \quad (10)$$

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:

$$w_i = (N_c - i)^2 \quad (11)$$

The value of z_i is defined as:

$$z_i = \begin{cases} 1 & \text{if } \text{nearest}(s_c, i) \in c \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

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 neighbours.

The Correct Classification Rate (CCR) is computed using the K -nearest neighbour (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 dataset. The recognition rate is computed as the average of all validations.

4 Experimental Results

The view-invariant gait analysis has been tested on real data from a subset of CASIA-B database [3] with subjects walking along straight lines with 6 different camera orientations (36° , 54° , 72° , 90° , 108° , 126°). The 90° view corresponds to the side view walking direction as shown in Figure 1. The video sequences have a spatial resolution and frame rate of 320x240 pixels and 25fps respectively with an approximate subject height of 90 pixels. Subjects are instructed to walk in three different scenarios; normal walking, wearing a coat and carrying a bag.

4.1 Performance Analysis of Gait Feature Extraction

The markerless algorithm has been applied to the video sequences, the limbs pose has been estimated frame by frame and the hip and knee angles have been extracted for each camera position and for each subject. Figure 1 shows an example of the limbs pose estimation for different camera positions for an example subject carrying a bag. The algorithm allows for estimation of the limbs pose also in such conditions and appears robust with respect to low-resolution, clothing and carrying conditions.

A quantitative validation of the proposed method has been obtained in a virtual environment (Poser 7[®]) with a humanoid walking for one gait cycle. The results reported in [5] for estimating the leg's angles have a mean value of 2.63 ± 2.61 deg and are particularly encouraging since they present same magnitude to the ones obtained with 3D markerless systems and 2D complex model based methods [11].

Figure 2(a) shows an example of the variations of hip angular motion during two gait cycle for the six different camera positions in the real experimental tests. Predictably, the angles trends, extracted in the image reference system, are influenced by the subject pose respect to the camera and they cannot be used directly for identification. For this reason, the view point correction algorithm is applied and the angle trends after the correction are shown in figure 2(b).

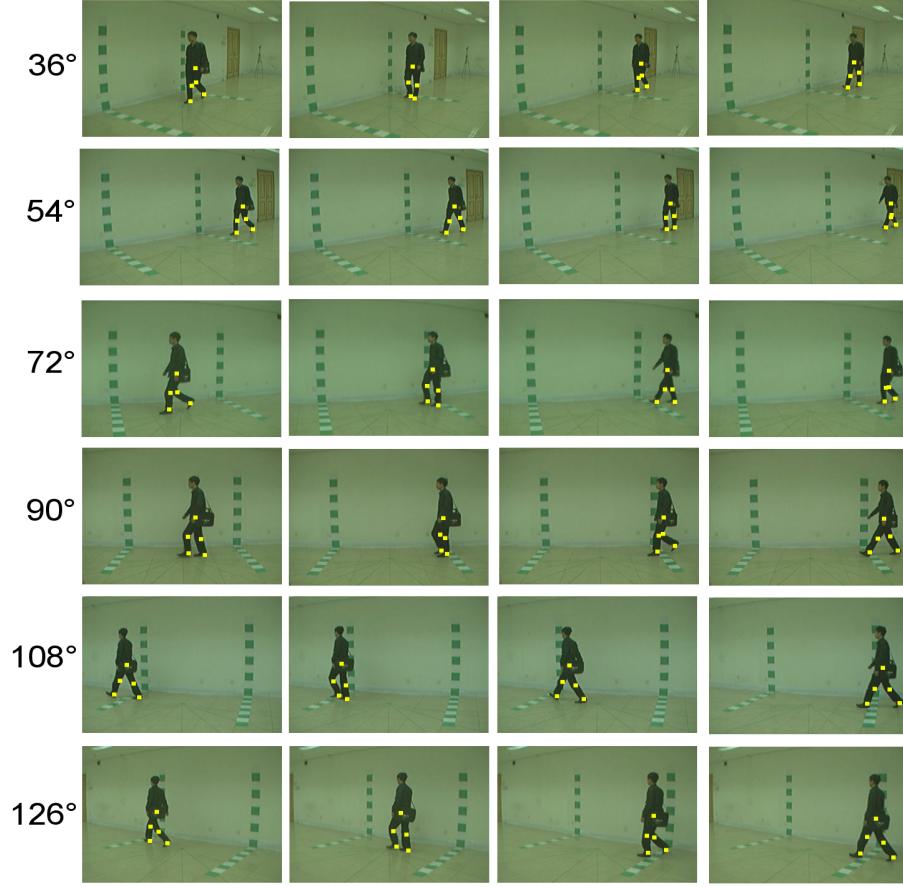


Fig. 1. Joints extraction in different viewpoints for subject carrying a bag.

4.2 Gait Recognition and Impact of Covariate Factors

In order to assess the performance of the proposed gait recognition algorithm from different viewpoints using a single uncalibrated camera, a set of 1037 video sequences with 20 different subjects recorded at 6 viewpoints are taken from the CASIA-B gait database. To investigate the effects of the viewpoint, an initial experiment is carried out to measure the recognition rate using the non-rectified gait data. The CCR is first computed for all the data combined together where a low CCR of 34% is observed based on leave-one-out cross validation.

A set of experiments are carried out to compute the recognition rates for every viewpoint separately after applying the view-rectification approach. This is done based on probing various datasets of different and similar viewpoints. Table 1 shows the variation of the CCRs with respect to the different viewpoints for the achieved results along with comparative results reported by Yu *et al.* [3]

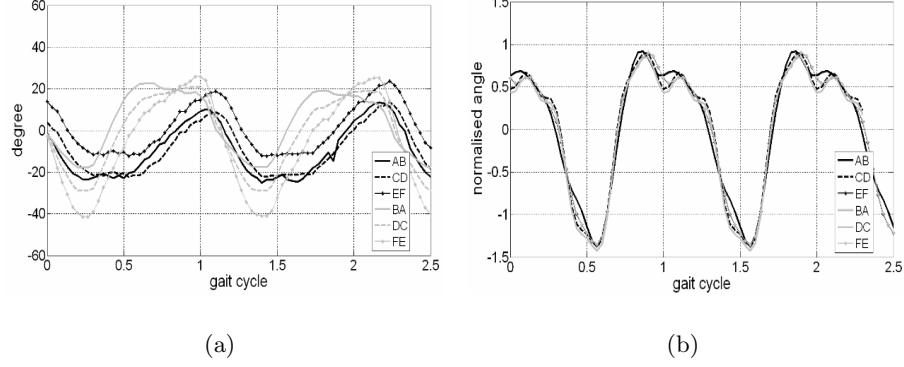


Fig. 2. Hip Angular Motion from different View-points: (a) Unrectified Data. (b) Rectified Angular Data.

in their silhouette-based approach applied on the CASIA-B dataset. The performance of gait recognition largely increases with an average CCR of 73.4% and better classification rates compared to the baseline silhouette-based approach. For both model and silhouette-based methods, the recognition rates along the diagonal for probing dataset against galleries of similar viewpoints, are observed to be higher with an average CCR of 80.8% and 30.15% for our method and the silhouette approach respectively. For the non-diagonal cases, the classification rates drop largely to an average of 9.6% for the silhouette-based due the changes of the silhouette shape when varying the viewpoint which affected the recognition performance. In contrast for the proposed model-based approach, a reported average CCR of 64.48% for probing datasets against probes of different viewpoints. 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 are which handles practical factors in human movement analysis.

Our Method (Rectified Data)							Yu <i>et al.</i> (Silhouette-Based)							
	Probe Angle							Probe Angle						
	36°	54°	72°	90°	108°	126°		36°	54°	72°	90°	108°	126°	
Gallery Angle	36°	67.8	60.8	58.9	50.1	48.0	40.8	36°	30.2	16.5	1.2	1.2	1.6	6.9
	54°	57.1	75.3	76.0	65.3	67.9	60.3	54°	10.1	30.6	5.6	4.4	7.7	14.1
	72°	52.3	63.3	83.4	81.5	79.0	72.7	72°	5.6	7.7	31.0	21.8	14.9	8.9
	90°	5.9	65.1	71.0	88.1	86.5	82.3	90°	4.0	6.0	20.6	32.7	16.5	6.0
	108°	44.7	61.0	68.9	79.6	86.6	81.9	108°	2.4	4.8	17.7	27.8	30.2	9.3
	126°	38.8	58.2	68.8	65.3	72.6	79.3	126°	1.6	4.4	10.1	10.1	18.5	26.2

Table 1. CCR (%) CASIA-B, Set A: 1) rectified data, 2) results of Yu *et al.* [3]

4.3 Covariate Analysis of Gait with Fixed View-point

To further describe the covariate effects, an experimental analysis was carried out on the SOTON Covariate database independently from the view-point rectification. A gallery dataset of 160 video sequences is taken from the SOTON gait database consisting of 20 different walking subjects with 8 sequences for every individual recorded without covariate effects. Further, a probe dataset of 440 video sequences for 10 subjects is collected from the Southampton Covariate Database. The covariate factors includes clothing, footwear, carrying conditions as well as walking speed. Based on the subset of features derived using the Feature Selection algorithm, we have achieved a high recognition rate of 95.75% for the value of $k = 5$ using the training covariate-free dataset. This is achieved using solely features describing purely the dynamics of the locomotion process.

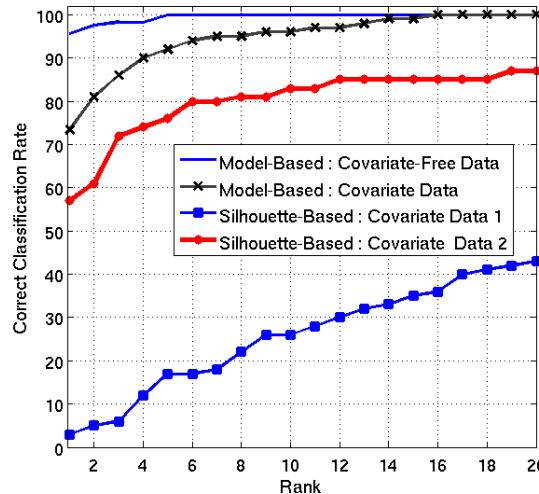


Fig. 3. The Cumulative Match Score Curves for the Classification Results.

Furthermore, we have probed 440 samples from the covariate dataset against the gallery database. A recognition rate of 73.4% is achieved for all the covariate factors which is higher when compared to the low recognition rates reported by Phillips *et al.* [2] using a silhouette-based method. The Cumulative Match Score curves showing the comparative results are shown in Figure (4.3). Phillips reported a CCR of 57% for Data (I) with load carriage and footwear covariates whilst a CCR of 3% is achieved for Data (II) with the following covariates : time, footwear, and clothing. Time has been shown [2] to play a major part in reducing recognition capability by gait. Using a silhouette based approach Veres[12] showed that this could be redressed by fusing those parts of the gait signature which are invariant with time.

5 Conclusions and Future Work

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 is described. Gait features are derived based on pose estimation of the joint positions of walking subjects. A novel reconstruction method is being employed to rectify and normalize gait features recorded from different viewpoint into the side-view plane and therefore exploit such data for recognition. The method is used to investigate of the effects of the covariate factors including clothing and carrying conditions for view-point independent gait recognition. Based on covariate-based probe datasets, a high recognition rate of 73.4% is achieved using the KNN classifier with $k = 5$. This suggests that people identification using dynamic gait features is still perceivable with better recognition rate even under the different covariate factors.

References

1. Nixon, M.S., Tan, T.N., Chellappa, R.: Human Identification Based on Gait (The Kluwer International Series on Biometrics). Springer-Verlag New York, Inc. Secaucus, NJ, USA (2005)
2. Sarkar, S., Phillips, P.J., Liu, Z., Vega, I.R., Grother, P., Bowyer, K.W.: The humanID Gait Challenge Problem: Data Sets, Performance, and Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27**(2) (2005) 162–177
3. Yu, S., Tan, D., Tan, T.: A Framework for Evaluating the Effect of View Angle, Clothing and Carrying Condition on Gait Recognition. *Proceedings of the 18th International Conference on Pattern Recognition* (2006) 441–444
4. Bouchrika, I., Nixon, M.: Exploratory factor analysis of gait recognition. In: *8th IEEE Int Conference on Automatic Face and Gesture Recognition*. (2008)
5. Goffredo, M., Seely, R.D., Carter, J.N., Nixon, M.S.: Markerless view independent gait analysis with self-camera calibration. In: *IEEE International Conference on Automatic Face and Gesture Recognition 2008*. (2008)
6. Bouchrika, I., Nixon, M.S.: Model-Based Feature Extraction for Gait Analysis and Recognition. In *Proceedings of Mirage: Computer Vision / Computer Graphics Collaboration Techniques and Applications* (2007) 150–160
7. Aguado, A.S.V., Nixon, M.S., Montiel, M.E.: Parameterizing Arbitrary Shapes via Fourier Descriptors for Evidence-Gathering Extraction. *Computer Vision and Image Understanding* **69**(2) (1998) 202–221
8. Chau, T.: A review of analytical techniques for gait data. Part 1: fuzzy, statistical and fractal methods. *Gait Posture* **13**(1) (2001) 49–66
9. Wagg, D.K., Nixon, M.S.: On Automated Model-Based Extraction and Analysis of Gait. In *Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition* (2004) 11–16
10. Somol, P., Pudil, P., Novovičová, J., Paclík, P.: Adaptive Floating Search Methods in Feature Selection. *Pattern Recognition Letters* **20**(11-13) (1999) 1157–1163
11. Agarwal, A., Triggs, B.: Recovering 3d human pose from monocular images. *IEEE TPAMI* **28** (2006) 44–58
12. Veres, G.V., Nixon, M.S., Middleton, L., Carter, J.N.: Fusion of Dynamic and Static Features for Gait Recognition over Time. In *Proceedings of 7th International Conference on Information Fusion* **2** (2005)