Model-Based Feature Extraction for Gait Analysis and Recognition

Imed Bouchrika and Mark S Nixon

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

Abstract. Human motion analysis has received a great attention from researchers in the last decade due to its potential use in different applications. We propose a new approach to extract human joints (vertex positions) using a model-based method. Motion templates describing the motion of the joints as derived by gait analysis, are parametrized using the elliptic Fourier descriptors. The heel strike data is exploited to reduce the dimensionality of the parametric models. People walk normal to the viewing plane, as major gait information is available in a sagittal view. The ankle, knee and hip joints are successfully extracted with high accuracy for indoor and outdoor data. In this way, we have established a baseline analysis which can be deployed in recognition, marker-less analysis and other areas. The experimental results confirmed the robustness of the proposed method to recognize walking subjects with a correct classification rate of %92.

Key words: Model-based feature extraction, motion analysis, gait, gait analysis.

1 Introduction

Much research in computer vision is directed into the analysis of articulated objects and more specifically, the analysis of human motion. This research is fuelled by the wide range of applications where human motion analysis can be deployed such as virtual reality, smart surveillance, human computer interfaces and athletic performance. A vision-based system for human motion analysis consists of three main phases: detection, tracking and perception. In the last phase, a high-level description is produced based on the features extracted during the previous phases from the temporal video stream. In fact, it has been revealed by psychological studies that the motion of human joints contains enough information to perceive the human motion.

Currently, the majority of systems used for motion analysis are marker-based and they are commercially available. Marker-based solutions rely primarily on markers or sensors attached at key locations of the human body. However, most applications such as visual surveillance require the deployment of an automated markerless vision system to extract the joints' trajectories. On the other hand,

automated extraction of the joints' positions is an extremely difficult task as non-rigid human motion encompasses a wide range of possible motion transformations due to its highly flexible structure and to self occlusion. Clothing type, segmentation errors and different viewpoints pose a significant challenge for accurate joint localization.

Since human motion analysis is one of the most active and challenging research topics in computer vision, many research studies have aimed to develop a system capable of overcoming the difficulties imposed by the extraction and tracking of human motion features. Various methods are surveyed by [1] and [2].Two approaches are being used for human motion anaylsis: model-based and non-model based methods. For the the first one, a priori shape model is established to match real images to this predefined model, and thereby extracting the corresponding features once the best match is obtained. Stick models and volumetric models [3] are the most commonly used methods. Akita [4] proposed a model consisting of six segments comprising of two arms, two legs, the torso and the head. Guo *et al* [5] represented the human body structure in the silhouette by a stick figure model which had ten sticks articulated with six joints. Rohr [6] proposed a volumetric model for the analysis of human motion, using 14 elliptical cylinders to model the human body. Recently, Karaulova *et al.* [7] have used the stick figure model to build a novel hierarchical model of human dynamics represented using hidden Markov models. The model-based approach is the most popular method being used for human motion analysis due to its advantages [8]. It can extract detailed and accurate motion data, as well as having the capability to cope well with occlusion and self-occlusion.

For the non-model based method, feature correspondence between successive frames is based upon prediction, velocity, shape, texture and colour. Shio *et al.* [9] proposed a method to describe the human body using moving blobs or 2D ribbons. The blobs are grouped based on the magnitude and the direction of the pixel velocity. Kurakake and Nevatia [10] worked on the extraction of joint locations by establishing correspondence between extracted blobs. Small motion between consecutive frames is the main assumption, whereby feature correspondence is conducted using various geometric constraints.

As there have been many vision approaches aimed to extract limbs, and a dearth of approaches specifically aimed to determine vertices, we propose a new method to extract human joints with better accuracy then blobs via incorporating priori knowledge to refine accuracy. Our new approach uses a model-based method for modelling human gait motion using elliptic Fourier descriptors, whereby the gait pattern is incorporated to establish a model used for tracking and feature correspondence. The proposed solution has capability to extract moving joints of human body with high accuracy for indoor data as well as outdoor data filmed in an unconstrained environments. Further, we assess the recognition capability using gait to demonstrate the potency of the approach described.

This paper is structured as follows: in the next section, we describe the method used for parameterizing arbitrary moving shapes using the elliptic Fourier

descriptors. Section 3 is devoted to the discussion of the approach adopted for the localization of human joints. Finally, the experimental results on a set of processed videos from the SOTON databases are drawn in section 4.

2 Arbitrary Shape Parametrization using Fourier Descriptors

A new model-based approach is proposed to extract the joints' trajectories of walking people. Although, the Fourier series is the most accurate way for modelling gait motion, most previous methods adopted simple models [11] to extract gait angular motion via evidence gathering using a few parameters. This is mainly due to complexity and computational cost. Grant *et al* [12] presented a new temporal evidence gathering method to extract arbitrary moving shapes. Fourier descriptors are used to parametrize the templates of moving shapes in a continuous form.

The Fourier theory has been used for the analysis of curves and boundaries of shapes for several years. The Fourier analysis provides a means for extracting features or descriptors from images which are useful characteristics for image understanding. These descriptors are defined by expanding the parametric representation of a curve in Fourier series. Let f be the function for the boundary of a given shape, the function f is represented using elliptic Fourier Descriptors [13, 14], where the Fourier series is based on a curve expressed by a complex parametric form as shown in equation (1):

$$f(t) = x(t) + iy(t) \tag{1}$$

where $t \in [0, 2\pi]$. $x(t)$ and $y(t)$ are approximated via the Fourier summation by n terms as shown in equation (2)

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} a_0 \\ b_0 \end{bmatrix} + \begin{bmatrix} X(t) \\ Y(t) \end{bmatrix} \tag{2}$$

such that a_0 and b_0 define the position of the shape's centre, and $X(t)$ and $Y(t)$ are computed as defined in equation (3) :

$$\begin{aligned} X(t) &= \sum_{k=1}^{n} a_{x_k} cos(kt) + b_{x_k} sin(kt) \\ Y(t) &= \sum_{k=1}^{n} a_{y_k} cos(kt) + b_{y_k} sin(kt) \end{aligned} \tag{3}$$

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 a Riemann summation [14]. For a representation invariant to rotation and scaling, we need to represent f in a parametrized form to cover all the possible graphs or shapes which can be derived by applying appearance transformation to the function f including rotation and scaling. Henceforth, the function f can be rewritten in the parametric form shown in equation (4):

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} a_0 \\ b_0 \end{bmatrix} + \begin{bmatrix} cos(\alpha) & -sin(\alpha) \\ sin(\alpha) & cos(\alpha) \end{bmatrix} \begin{bmatrix} X(t) * s_x \\ Y(t) * s_y \end{bmatrix} \tag{4}$$

where α is the rotation angle, s_x and s_y are the scaling factors across the horizontal and vertical axes respectively. The last equation (4) can be written concisely in its complex form as defined in equation (5):

$$\begin{cases} f = T + R_\alpha \left(s_x X(t) + s_y Y(t) i \right) \\ T = a_0 + b_0 i \\ R_\alpha = \cos(\alpha) + \sin(\alpha) i \end{cases} \tag{5}$$

Based on the final parametric format of f shown in equation (5), any shape can be represented using five parameters which are: a_0, b_0, α, s_x and s_y. $X(t)$ and $Y(t)$ are pre-computed using equation (3) from the original shape. In fact, the number of free parameters needed for the Hough Transform is totally independent of the complexity of shape which is defined using the elliptic Fourier Descriptors, as the defined parameters are related to the appearance transformations which define all the shapes that can be derived form the original shape.

To use the Hough transform with these templates represented via the parametric form described in equation (5), a five-dimensional space is required. Thus, the algorithm would be computationally intensive and infeasible to implement. In spite of the fact that some methods were proposed to reduce the computational requirements of the Hough Transform [15, 16], the computational load of these methods does not meet the requirements of most applications [16]. Alternatively, the heel strike data could be incorporated to reduce the complexity of the parameter space and therefore, dramatically reduce the computational cost as done here.

3 Extraction of the Anatomical Landmarks

3.1 Heel Strike Extraction

The detection of the human gait period can provide important information to determine the positions of the human joints. Cutler *et al* [17] proposed a real time method for measuring periodicity for periodic motion based on self-similarity. Instead, the heel strikes of the subject can provide an accurate measure for gait periodicity as well as the gait stride and step length. Moreover, the extraction of heel strikes can be used as a strong cue to distinguish walking people from other moving objects in the scene [18].

During the strike phase, the foot of the striking leg stays at the same position for half a gait cycle, whilst the rest of the human body moves forward. Therefore, if we use a low-level feature extraction method (edges or corners), then a dense region will be accumulated at the heel strike regions. Since the primary aim of this research is the perception of human motion, we have chosen to use corners instead of edges, as they maintain enough information to perceive the human motion, in contrast to edges which may cause ambiguity in the extraction process due to the excess data they may contain. Furthermore, a robust vision system based on corner detection can work for low-resolution applications. We have

applied the Harris corner detector on every frame t from the video sequence and then accumulated all the corners into one image using equation (6):

$$C_i = \sum_{t=1}^{N} H(I_t) \tag{6}$$

where H is the output of the Harris corner detector, I_t is original image at frame t. Because the striking foot is stabilized for half a gait cycle, as result, a dense area of corners is detected in the region where the leg strikes the ground. In order to locate these areas, we have estimated a measure for density of proximity. The value of proximity at point p is dependent on the number of corners within the region R_p and their corresponding distances from p. R_p is assumed to be a square area with centre p, and radius of r that is determined as the ratio of total image points to the total of corners in C_i which is about 10. We have first computed proximity value d_p of corners for all regions R_p in C_i using equation (7). This is an iterative process starting from a radius r. The process then iterates to accumulate proximity values of corners for point p.

$$\begin{cases} d_p^r = \frac{N_r}{r} \\ d_p^i = d_p^{i+1} + \frac{N_i}{i} \end{cases} \tag{7}$$

where d_p^i is the proximity value for rings of radius i away from the centre p, and N_i is the number of corners which are of distance i from the centre, rings are single pixel wide. Afterwards, we accumulate all the densities for the subregions R_p for all points p into one image to produce the corners proximity image using (8).

$$D = \sum_{x=0}^{X} \sum_{y=0}^{Y} shift(d_p) \tag{8}$$

where X and Y are the width and height of the image respectively. d_p is the corners proximity value for region R_p. The $shift$ function places the proximity value d_p on a blank image of size $X \times Y$ at the position p. An output of the corner proximity for an example image is shown in Figure (1). The input image contains points spread all over the image with a number dense regions. The resulting image has darker areas which correspond to the crowded regions in the input image.

Figure (2) shows the corner proximity images for two walking subjects being captured in different environments. The first subject is walking in the sagittal plane near the camera, whilst the second subject is recorded in an oblique view walking away from the camera. A similar algorithm to [19] is used to derive the positions of the peaks as local maxima.

3.2 Moving Joints Extraction

In order to recognize people by their gait, the model-based method is used for the localization of the joints for walking people. The evidence gathering process
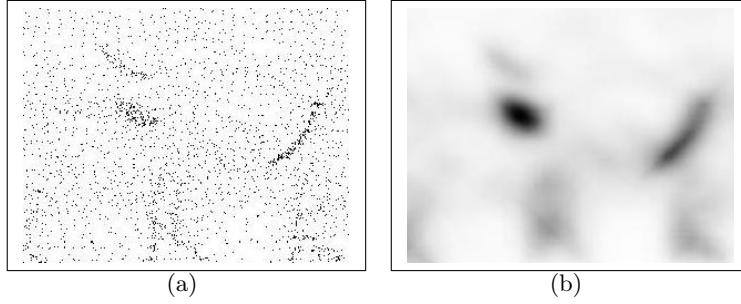
(a)   (b)

Fig. 1. Example Results for the Corner Proximity Measure: (a) Input Image, (b) Corner Proximity Image.

is the most common method used for the extraction of articulated objects. It is usually used in conjunction with the Hough Transform consisting of two phases: i) global pattern extraction, and ii) local feature extraction. The objective of the global extraction is to find the best motion pattern based on the predefined model i.e. prior knowledge from the whole video sequence. The motion model is defined in a parametric form using the elliptic Fourier Descriptors. The Hough Transform is used to determine the optimum parameters through the matching process of feature points (corners) across the whole sequence to the parametric function, and increase votes in the accumulator space accordingly. The second stage of the evidence gathering is a frame-by-frame search aimed to extract the precise position of the joint. The local extraction is guided by the pattern extracted at the first stage.

Model templates which describe joints' motion are derived as the mean values through the analysis of manually labelled data of 30 video sequences. Figure (3) shows the joint motion paths between two successive strikes of the same leg for the ankle, knee and hip. Because only closed and continuous contours can be represented by the elliptic Fourier descriptors, we have converted the model template into a closed cyclic contour by looping back along the graph. The voting process for the evidence gathering method is applied on corners images which are obtained by appliying the Harris corner detector on every frame. Anatomical knowledge [20] of the body segments properties is used to limit the search area.

For the extraction of the ankle motion path, the five-dimensional parametric equation (5) is rewritten in terms of only one parameter s_y as described in equation (9). Because the heel strike points are known and lie on the motion graph of the ankle as both ends of the graph, henceforth, this knowledge is used to deduce the two parameters s_x and α as the distance and angle between the two strikes respectively.

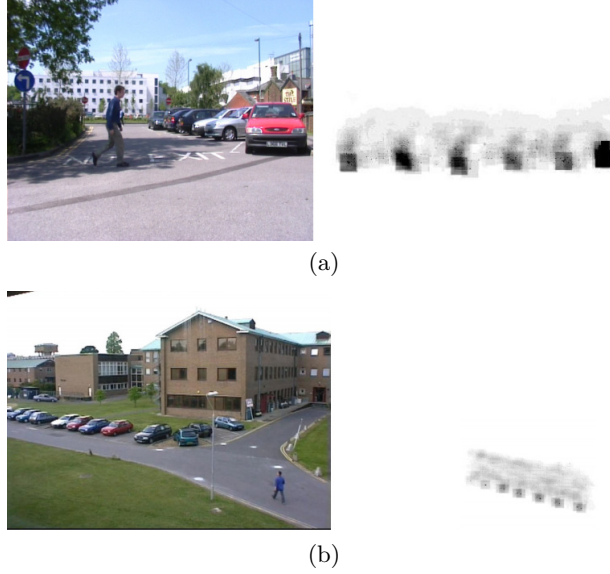$$s_y = \frac{(y - y_{s_n}) - (X(t) - x_e)s_x \sin(\alpha)}{(Y(t) - y_e)\cos(\alpha)} \tag{9}$$

(a)

(b)

Fig. 2. Heel Strike Extraction using the Proximity Measure: (a) Sagittal Indoor View, (b) Oblique Outdoor View.

where (x, y), (x_e, y_e) and (x_{s_n}, y_{s_n}) are the coordinates of the locus point, left end point of the model template and the nth strike respectively. The translation parameters are then computed after determining the best value of s_y using the following equation (10):

$$T = s_n - (\cos(\alpha) + i\sin(\alpha))(x_e s_x + y_e s_y i) \tag{10}$$

where i is the complex number. In the same way, the dimensionality of the parametric equation is reduced to three parameters for the global search of the hip and knee joints.

After obtaining the global motion pattern for the joints between the strikes s_n and s_{n+2}, a local search is performed within every frame to find the precise position of the joints. By gait analysis, people more or less have the same horizontal velocity pattern for the ankle, hip and knee displacement when walking. Therefore, a velocity model is produced and used to estimate the spread of x values through the temporal data. Let N the number of frames between the two strikes s_n and s_{n+2}. The x values for the joints trajectories can be estimated as expressed by the following equation:

$$x_f = s_x \times V\left(\frac{f}{N}\right) + x_{si} \tag{11}$$

where f is the frame number, x_s is the x coordinate of the heel strike s_n and V is the horizontal velocity model function for the joint (ankle, hip or knee). After
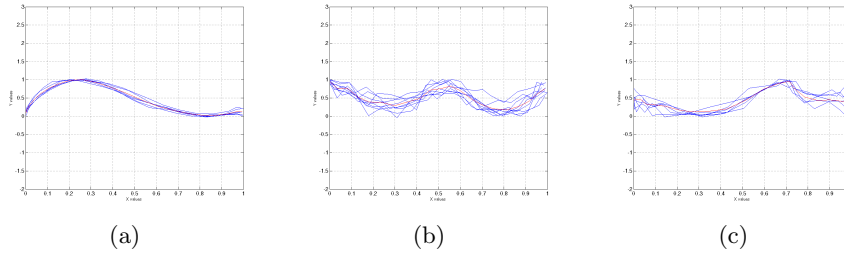
**Fig. 3.** Motion Analysis of the Joints: (a) Ankle, (b) Hip , (c)Knee

getting the x_f values, then we use the motion graph derived from the global search to obtain the y coordinates of the joints.

4   Experimental Results

To demonstrate the efficacy of this approach, we have run the algorithm on a set of 120 different subjects from the SOTON database [21]. 100 subjects of the test database were filmed in an indoor environment with controlled conditions. The remaining 10 subjects were filmed in outdoor environment. Subjects walked from left to right normal to the viewing plane.

For the detection of the heel strike points, the proposed method extracted successfully %99.2 of the strikes from a total of 514 strikes. The mean error for the positions of 65 strikes extracted by the algorithm compared to strikes manually labelled is %0.52 of the person height. The error is measured by Euclidean distance between the two strikes normalized to a percentage of the person's height as this is the normal basis for normalization. We have extracted the joints for the ankles, knees and hip as shown in Figure (4). The mean error for the positions of the extracted joints compared with manual data of 10 subjects manually labelled is %1.36 of the height of the subject. The algorithm is tested on a subject wearing Indian clothes which covered the legs. The joints positions are extracted successfully as shown in Figure 4(b) which reveals the potentials of this approach to handle occlusion. Figure (5) shows the relative angle for both the hip and knee computed from the extracted joints of 10 subjects. The graphs show the results obtained via this approach are consistent with biomechanical data by Winter [22] shown in bold in Figure (5).

To evaluate the recognition performance of our algorithm, we have fused both dynamic and static gait features to yield a feature vector consisting of 48 features. Static features include the body height, stride and heights of different body parts whilst dynamic features are mainly the phase-weighted magnitudes of the Fourier frequencies for the hip and knee angular motions. The gait signature is derived using the adaptive forward floating search algorithm via selecting the features with higher discriminability values. The gait signatures for every sequence are projected into the canonical space whereby the k-nearest neighbor
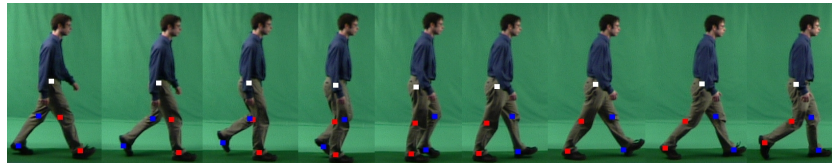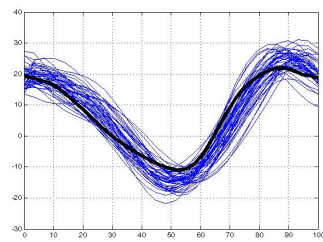
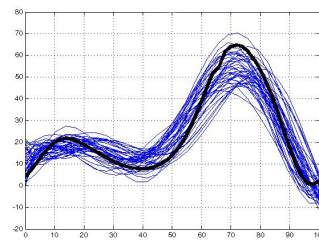(a) Subject : 009a020s00R.

(b) Subject : 012a031s00R.

(c) Subject : 012a031s00R.

Fig. 4. Joints Extraction for Indoor Data.

rule is used for classification. The system achieved a correct classification rate of %92 using the leave-one-out validation rule. This is shown in Figure (6) with only 10 of the 20 subjects being projected.

(a)   (b)

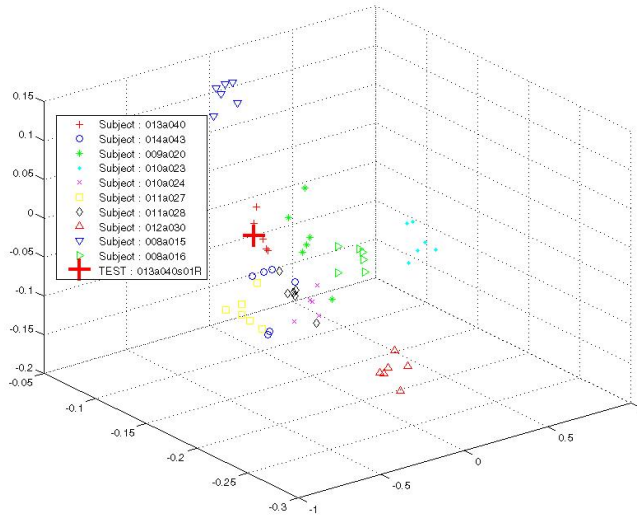**Fig. 5.** Gait Angular Motion during one Gait Cycle: (a) Hip, (b) Knee

Fig. 6. Canonical Space Projection

5 Conclusions and Future Work

We propose a model-based method to extract moving joints via evidence gathering technique. Spatial model templates for human motion are derived from the analysis of gait data collected from manual labelling. Model templates are represented in a parametric form based on elliptic Fourier descriptors. Gait knowledge is exploited via heel strike extraction to reduce the the parameter space dimensionality and therefore reduce the computational load of the Hough Transform being used in the extraction process. The described method is proved to to work for both indoor and outdoor environments with potential to localize joint positions with better accuracy. Furthermore, the recognition potency is evaluated using the data extracted via the described model-based method. The proposed solution has achieved a classification rate of %92 for people recognition. The model-based is suited to more generalized deplyment and this will be the focus for future work.

References

1. Wang, L.A., Hu, W.M., Tan, T.N.: Recent developments in human motion analysis. Pattern Recognition **36**(3) (2003) pp. 585–601
2. Aggarwal, J.K., Cai, Q., Liao, W., Sabata, B.: Nonrigid motion analysis: Articulated and elastic motion. CVIU **70**(2) (1998) pp. 142–156

3. Yoo, J.H., Nixon, M.S., Harris, C.J.: Extraction and description of moving human body by periodic motion analysis. Proc. ISCA 17th International Conference on Computers and Their Applications (2002) pp. 110–113

4. Akita, K.: Image sequence analysis of real world human motion. Pattern Recognition **17**(1) (1984) pp. 73–83

5. Guo, Y., Xu, G., Tsuji, S.: Understanding human motion patterns. Proc. the 12th IAPR ICPR. Volume 2. (1994) pp. 325–329

6. Rohr, K.: Towards model-based recognition of human movements in image sequences. CVGIP: IU **74**(1) (1994) pp. 94–115

7. Karaulova, I.A., Hall, P.M., Marshall, A.D.: A hierarchical model of dynamics for tracking people with a single video camera. Proc of the 11th BMVA (2000) pp. 262–352

8. Huazhong, N., Tan, T., Wang, L., Hu, W.: People tracking based on motion model and motion constraints with automatic initialization. Pattern Recognition **37**(7) (2004) pp. 1423–1440

9. Shio, A., Sklansky, J.: Segmentation of people in motion. In: Proc of the IEEE Workshop on Visual Motion. Volume 2. (1991) pp. 325–332

10. Kurakake, S., Nevatia, R.: Description and tracking of moving articulated objects. In: Proc. 11th IAPR ICPR. Volume 1. (1992) pp. 491–495

11. Cunado, D., Nixon, M., Carter, J.: Automatic Extraction and Description of Human Gait Models for Recognition Purposes. CVIU **90**(1) (2003) 1–41

12. Grant, M.G., Nixon, M.S., Lewis, P.H.: Extracting moving shapes by evidence gathering. Pattern Recognition **35**(5) (2002) 1099–1114

13. Granlund, G.H.: Fourier preprocessing for hand print character recognition. IEEE T-Comp **21** (1972) pp. 195–201

14. Aguado, A.S., Nixon, M.S., Montiel, M.E.: Parameterising arbitrary shapes via fourier descriptors for evidence-gathering extraction. CVGIP: IU **2** (1998) pp. 547–51

15. Leavers, V.F.: Which Hough transform? CVGIP: IU **58**(2) (1993) 250–264

16. Aguado, A., Nixon, M., Montiel, M.: Parameterizing Arbitrary Shapes via Fourier Descriptors for Evidence-Gathering Extraction. CVIU **69**(2) (1998) 202–221

17. Cutler, R., Davis, L.S.: Robust real-time periodic motion detection, analysis, and applications. IEEE TPAMI **22**(8) (2003) pp. 781–796

18. Bouchrika, I., Nixon, M.S.: People Detection and Recognition using Gait for Automated Visual Surveillance. IEE International Symposium on Imaging for Crime Detection and Prevention (2006)

19. Fujiyoshi, H., Lipton, A.J., Kanade, T.: Real-time human motion analysis by image skeletonization. IEICE Trans on Information and System (2004) 113–120

20. Pheasant, S.T.: Body sapce: Anthropometry, Ergonomics and Design. Taylor & Francis (1988)

21. Shutler, J.D., Grant, M.G., Nixon, M.S., Carter, J.N.: On a large sequence-based human gait database. In: Proc of Recent Advances in Soft Computing, Nottingham, UK (2002) pp. 66–71

22. Winter, D.A.: The Biomechanics and Motor Control of Human Movement. second edn. John Wiley & Sons (1990)