Markerless Feature Extraction for Gait Analysis

Imed Bouchrika and Mark S. Nixon
Department of Electronics and Computer Science
University of Southampton
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. The gait pattern is incorporated to aid the extraction process, where model templates are established through analysis of gait motion. People walk normal to the viewing plane, as major gait information is available in a sagittal view. Gait periodicity and other parameters are estimated by finding the heel strikes. 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.*

Keywords: Human motion analysis, gait analysis, markerless feature extraction.

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. This is mainly due to their accuracy. 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 [26, 10]. Clothing type, segmentation errors and different viewpoints pose a significant challenge for accurate joint localization.

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 in both indoor and outdoor environments.

1.1 Related Work

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 [23] and [1].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 [26] are the most commonly used methods. Akita [3] proposed a model consisting of six segments comprising of two arms, two legs, the torso and the head. Guo *et al* [13] represented the human body structure in the silhouette by a stick figure model which had ten sticks articulated with six joints. Rohr [20] proposed a volumetric model for the analysis of human motion, using 14 elliptical cylinders to model the human body. Recently, Karaulova *et al.* [16] 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 [14]. 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.* [21] 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 [18] 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.

1.2 System Overview

The system proposed in this paper consists of three stages as outlined in Figure (1). In the first stage, walking subjects are detected using background subtraction. The approach we used for the segmentation of moving objects, is the adaptive background subtraction proposed by Stauffer and Crimson. It is assumed only one single moving subject in the scene. The heel strikes are derived in the next stage after applying the Harris corner operator. Finally, the evidence gathering algorithm is applied to locate the joint positions.

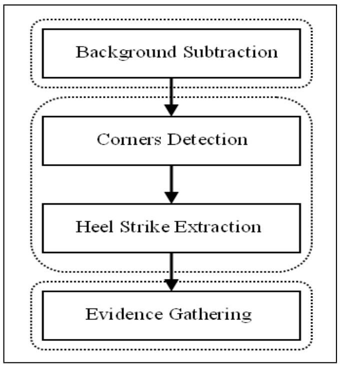


Figure 1: System Overview

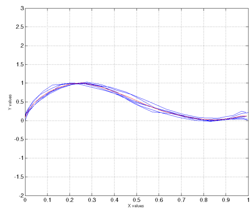
2 Extraction of the Anatomical Landmarks

2.1 Human Motion Analysis

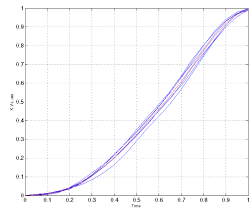
The motion of the human body is known as a form of non-rigid and articulated motion [1], and therefore detection and tracking the right information becomes an extremely difficult task as non-rigid motion encompasses a wide range of possible motion transformations due to the highly flexible structure and self occlusion. During walking and running, people have the same global gait motion pattern. Therefore, gait motion can be considered as an ideal starting point for motion analysis due to its global nature and since it is periodic.

Psychological studies carried out by Johansson [15] showed that people are able to perceive human motion from Moving Light Displays (MLD). An MLD is a two-dimensional video of a collection of bright dots attached to the human body taken against a dark background where only the bright dots are visible in the scene. An observer can recognise different types of human motion such as walking, jumping, dancing and so on. Moreover, the observer can make a judgment of the gender of the person [17], and even further identify the person if he or she is already familiar with his or her gait [11]. Although the different parts of the human body are not seen in the MLD, and no links exist between the bright dots to show the structure, the human can recover the full structure of the moving object. Thereby, the motion of the joints contains enough information to perceive human motion [4], [8].

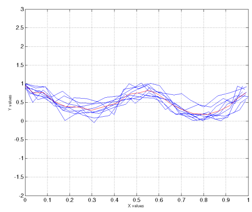
To analyse the human motion, the joint positions in 30 video sequences with people walking normal to the viewing plane of the camera, have been manually labelled. The videos are taken from the SOTON database. In each frame of the video sequence, the position of the right and left ankles, the right and left knees and the hip were labelled. The data for the ankle between to consecutive heel strikes of the same leg are normalized as shown in Figure 2(a). Whilst, we have normalized the data for the knee and hip extracted between two consecutive stances of the same leg as shown in Figures 2(c) and 2(e). The corresponding horizontal displacement for each joint is plotted against the motion graph of the joint in Figure (2).



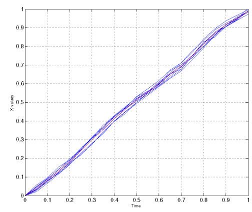
(a) Ankle Motion Graph.



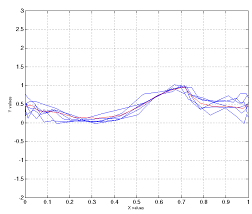
(b) Horizontal Ankle Displacement



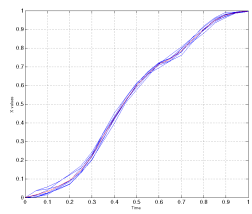
(c) Hip Motion Graph.



(d) Horizontal Hip Displacement.



(e) Knee Motion Graph.



(f) Horizontal Knee Displacement.

Figure 2: Motion Analysis of the Joints.

It can be observed that people have more or less the same ankle motion pattern. Another graph 2(b) is plotted showing the horizontal displacement of the ankle, where it is noted that the graphs for all subjects nearly coincide, leading to the suggestion that for a normalized data set, subjects move their ankles forward with the same velocity. Figures 2(c) and 2(e) show the hip and knee motions respectively for the normalized extracted data. In contrast to the smooth graphs of the ankle, there is noise in the data for the hip and knee due the difficulties encountered during the manual labelling. Nevertheless, it can be observed that walking people have the same global pattern for the hip and knee motions. The horizontal displacement for the hip and knee are shown in Figures 2(d) and 2(f). The hip forward velocity is decimated to be constant for all subjects, in contrast to the knee velocity, which varies for all subjects. However, with the data normalized, people have more or less the same knee horizontal velocity.

2.2 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* [7] 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 [5].

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 as shown in Figure (3). 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.

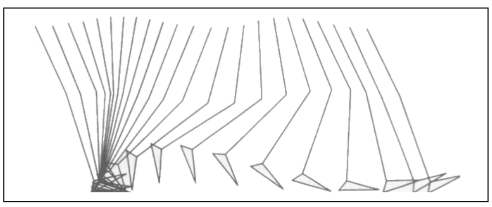


Figure 3: Foot Displacement during one Gait Cycle [24]

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 (1):

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

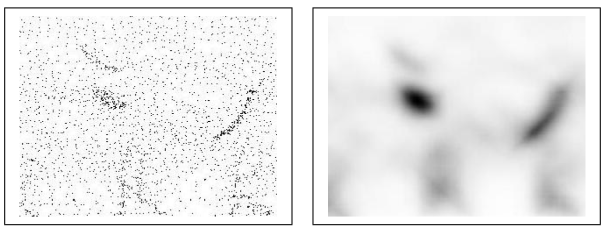
Where H is the output of the Harris corner detector, I_t is original image at frame t. Because the striking foot is stabilized for half a gait cycle, as result, a dense area of corners is detected in the region where the leg strikes the ground. In order to locate these areas, we have estimated a measure 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 (2). This is an iterative process starting from a radius r. The process then iterates to accumulate proximity values of corners for point p.

$$\begin{cases} d_p^r = \frac{N_r}{r} \\ d_p^i = d_p^{i+1} + \frac{N_i}{i} \end{cases} \quad (2)$$

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 (3).

$$D = \sum_{x=0}^{X} \sum_{y=0}^{Y} shift(d_p) \quad (3)$$

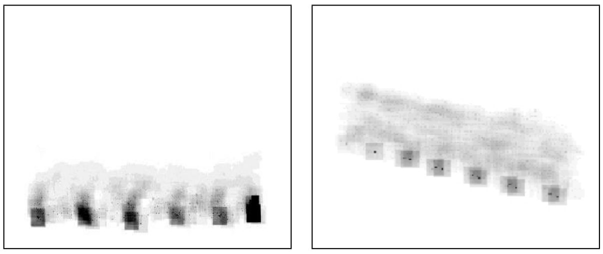
where X and Y are the width and height of the image respectively. d_p is the corners proximity value for region R_p. The $shift$ function places the proximity value d_p on a blank image of size $X \times Y$ at the position p. An output of the corner proximity for an example image is shown in Figure (4). 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.



(a) (b)

Figure 4: Example Results for the Corner Proximity Measure: (a) Input Image, (b) Corner Proximity Image.

Figure (5) 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 [9] is used to derive the positions of the peaks as local maxima.



(a) (b)

Figure 5: Heel Strike Extraction using the Proximity Measure: (a) Sagittal Indoor View, (b) Oblique Outdoor View.

2.3 Moving Joints Extraction

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 [6] to extract gait angular motion via evidence gathering using a few parameters. This is mainly due to complexity and computational cost. In our method, human gait is modelled using the Fourier series. The heel strike data were used to reduce the number of parameters and therefore reduce significantly the computational cost. Model templates which describe joints' motions are constructed from the analysis of manually labelled data.

The mean patterns for gait motion are represented using elliptic Fourier Descriptors [2]. 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 the motion models, the function f is represented using elliptic Fourier Descriptors [12, 2], where the Fourier series is based on a curve expressed by a complex parametric form as shown in equation (4):

$$f(t) = x(t) + jy(t) \tag{4}$$

where $x(t)$ and $y(t)$ are approximated via the Fourier summation by n terms as shown in equation (5) :

$$\begin{aligned} x(t) &= \sum_{k=1}^{n} a_{x_k} cos(kt) + b_{x_k} sin(kt) \\ y(t) &= \sum_{k=1}^{n} a_{y_k} cos(kt) + b_{y_k} sin(kt) \end{aligned} \tag{5}$$

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 [2]. In order to obtain a flexible motion model sufficient to describe gait motion, spatial model templates are created via representing f in a parametrized form by applying appearance transformations (rotation, scaling and translation). A spatial model template M describing gait motion is described in equation (6):

$$\begin{cases} M = T + R_\alpha \left(s_x x(t) + s_y y(t) i \right) \\ T = a_0 + b_0 i \\ R_\alpha = \cos(\alpha) + \sin(\alpha) i \end{cases} \tag{6}$$

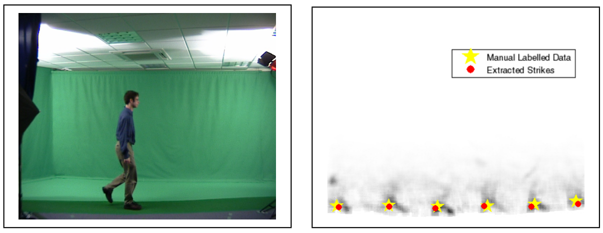
where T is the translation transform whose vector is (a_0, b_0). R is the rotation transform of angle α. s_x and s_y are the scaling factors across the horizontal and vertical axes respectively.

The evidence gathering process [6] is usually used in conjunction with the Hough Transform consisting of two phases: i) global pattern extraction, and ii) local feature extraction. The aim of the global extraction is to find the best motion pattern based on the predefined model represented using in a parametric form based on the elliptic Fourier Descriptors. The Hough Transform is used as a first stage to extract the spatial motion path of the joints using model templates. Because a 5-dimensional accumulator is needed to to store the votes for the set of parameters $< a_0, b_0, \alpha, s_x, s_y >$, the algorithm would be computationally intensive and infeasible to implement. In spite of the fact that some methods were proposed to reduce the processing time of the Hough Transform [19, 2], the computational load of these methods does not meet the requirements of most applications [2]. Alternatively, the heel strike data could be incorporated to reduce the complexity of the parameter space and therefore, dramatically reduce the computational cost. The search for the ankle motion model is reduced to only one parameter s_y, while it is reduced to two parameters b_0 and s_y for the case of the hip and knee motion models.

The second stage of the process is to apply a local search within every frame to determine the position of the joints. The local search is guided by the motion of the joints extracted in the first stage.

3 Experimental Results

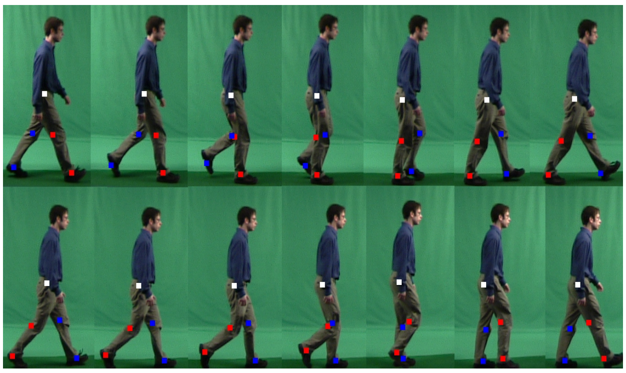
To demonstrate the efficacy of this approach, we have run the algorithm on a set of 100 different subjects from the SOTON database [22]. all subjects are filmed in an indoor environment with controlled conditions. subjects walked from left to right normal to the viewing plane. From a total of 514 strikes, the algorithm extracted successfully 510 strikes with only four strikes being missed. 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 normalized to a percentage of the person's height. Figure (6) shows the results of heel strike extraction by the described method compared with the data labelled manually for one video sequence and it can be observed that the match is indeed close.



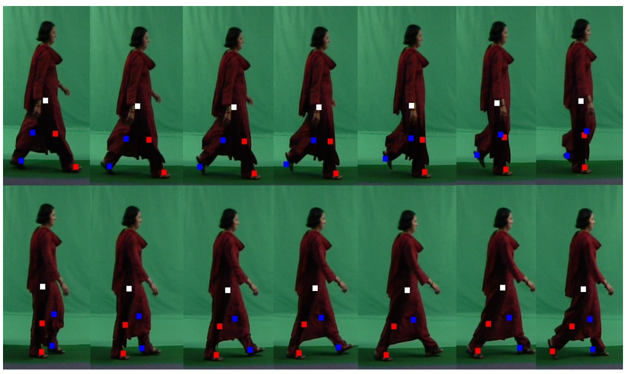
(a) (b)

Figure 6: Experimental Results for Heel Strikes Extraction: (a) Walking subject. (b) Extracted strikes compared with data manually labelled

We have extracted the joints for the ankles, knees and hip as shown in Figure (7). 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 7(b) which reveals the potentials of this approach to handle occlusion.



(a) Subject : 009a020s00R.

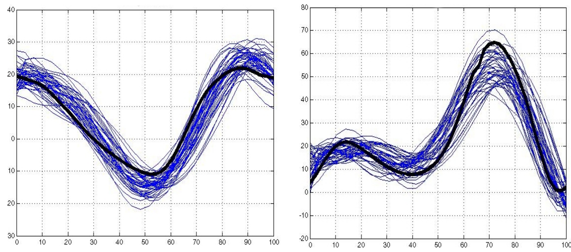


(b) Subject : 012a031s00R.

Figure 7: Joints Extraction for Indoor Data.

Figure (8) 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 the biomechanical data by Winter [25] shown in bold in Figure (8).

We have conducted further experiments in outdoor environment to confirm the robustness of the proposed method. The algorithm is tested on outdoor data containing 20 subjects from the SOTON database. The heel strikes are extracted successfully. The mean error for the positions of the joints extracted is estimated to %2 of the height of the person. Figure (9) shows the extraction results of the joints in outdoor data.



(a) (b)

Figure 8: Gait Angular Motion during one Gait Cycle: (a) Hip, (b) Knee

Figure 9: Joints Extraction for Outdoor Data.

4 Conclusions

We have proposed a new method to extract the positions human joints using a model-based method. The gait pattern is deployed to aid the extraction process, where model templates are established through the analysis of gait motion. Gait periodicity and other parameters are estimated through the finding of the heel strikes. The joints of the ankle, knee and hip are successfully extracted with high accuracy for indoor and outdoor data.

References

[1] J. K. Aggarwal, Q. Cai, W. Liao, and B. Sabata. Nonrigid motion analysis: Articulated and elastic motion. *Computer Vision and Image Understanding*, Vol 70, No. 2, pp. 142–156, 1998.

[2] A. S. Aguado, M. S. Nixon, and M. E. Montiel. Parameterizing Arbitrary Shapes via Fourier Descriptors for Evidence-Gathering Extraction. *Computer Vision*

and Image Understanding, Vol 69, No. 2, pp. 202–221, 1998.

[3] K. Akita. Image sequence analysis of real world human motion. *Pattern Recognition*, Vol 17, No. 1, pp. 73–83, 1984.

[4] G. P. Bingham, R. C. Schmidt, and L. D. Rosenblum. Dynamics and the orientation of kinematic forms in visual event recognition. *Journal of Experimental Psychology: Human Perception and Performance*, Vol 21, No. 6, pp. 1473-1493, 1991.

[5] I. Bouchrika and M. S. Nixon. People Detection and Recognition using Gait for Automated Visual Surveillance. London, UK, June 2006.

[6] D. Cunado, MS Nixon, and JN Carter. Automatic Extraction and Description of Human Gait Models for Recognition Purposes. *Computer Vision and Image Understanding*, Vol 90, No. 1, pp. 1–41, 2003.

[7] R. Cutler and L. S. Davis. Robust real-time periodic motion detection, analysis, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol 22, No. 8, pp. 781–796, 2003.

[8] W. H. Dittrich. Action categories and the perception of biological motion. *Perception*, Vol 22, pp. 15-22, 1993.

[9] H. Fujiyoshi, A. J. Lipton, and T. Kanade. Real-time human motion analysis by image skeletonization. *IEICE Trans on Information and System*, pp. 113–120, 2004.

[10] D. M. Gavrila. The visual analysis of human movement: A survey. *Computer Vision and Image Understanding*, Vol 73, No. 1, pp. 82–98, 1999.

[11] N. H. Goddard. *The perception of articulated motion: Recognizing moving light displays*. PhD thesis, University of Rochester, 1992.

[12] G. H. Granlund. Fourier preprocessing for hand print character recognition. *IEEE T-Comp*, Vol 21, pp. 195–201, 1972.

[13] Y. Guo, G. Xu, and S. Tsuji. Understanding human motion patterns. In *Proceedings of the 12th IAPR International Conference on Pattern Recognition*, Vol 2, pp. 325–329, 1994.

[14] N. Huazhong, T. Tan, L. Wang, and W. Hu. People tracking based on motion model and motion constraints with automatic initialization. *Pattern Recognition*, Vol 37, No. 7, pp. 1423–1440, 2004.

[15] G. Johansson. Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics*, Vol 14, pp. 201-211, 1973.

[16] I. A. Karaulova, P. M. Hall, and A. D. Marshall. A hierarchical model of dynamics for tracking people with a single video camera. In *Proceedings of the 11th British Machine Vision Conference*, pp. 262–352, Bristol, Septemeber 2000.

[17] L. T. Kozlowski and J. E. Cutting. Recognizing the gender of walkers from point-lights mounted on ankles: Some second thoughts. *Perception & Psychophysics*, Vol 23, pp. 459, 1978.

[18] S. Kurakake and R. Nevatia. Description and tracking of moving articulated objects. In *Proceedings. 11th IAPR International Conference on Pattern Recognition*, Vol 1, pp. 491–495, Octobor 1992.

[19] V. F. Leavers. Which Hough transform? *CVGIP: Image Understanding*, Vol 58, No. 2, pp. 250–264, 1993.

[20] K. Rohr. Towards model-based recognition of human movements in image sequences. *CVGIP: Image Understanding*, Vol 74, No. 1, pp. 94–115, 1994.

[21] A. Shio and J. Sklansky. Segmentation of people in motion. In *Proceedings of the IEEE Workshop on Visual Motion*, Vol 2, pp. 325–332, Octobor 1991.

[22] J. D. Shutler, M. G. Grant, M. S. Nixon, and J. N. Carter. On a large sequence-based human gait database. In *Proceedings of Recent Advances in Soft Computing*, pp. 66–71, Nottingham, UK, 2002.

[23] L. A. Wang, W. M. Hu, and T. N. Tan. Recent developments in human motion analysis. *Pattern Recognition*, Vol 36, No. 3, pp. 585–601, 2003.

[24] M. Whittle. *Gait Analysis: An Introduction*. Butterworth-Heinemann, 2002.

[25] D. A. Winter. *The Biomechanics and Motor Control of Human Movement*. John Wiley & Sons, second edition, 1990.

[26] J. H. Yoo, M. S. Nixon, and C. J. Harris. Extraction and description of moving human body by periodic motion analysis. *Proceedings of ISCA 17th International Conference on Computers and Their Applications*, pp. 110–113, 2002.