A Layered Deformable Model for Gait Analysis

Haiping Lu, K.N. Plataniotis and A.N. Venetsanopoulos
Bell Canada Multimedia Laboratory
The Edward S. Rogers Sr. Department of Electrical and Computer Engineering
University of Toronto, M5S 3G4, Canada
{haiping, kostas, anv}@dsp.toronto.edu

Abstract

In this paper, a layered deformable model (LDM) is proposed for human body pose recovery in gait analysis. This model is inspired by the manually labeled silhouettes in [6] and it is designed to closely match them. For fronto-parallel gait, the introduced LDM model defines the body part widths and lengths, the position and the joint angles of human body using 22 parameters. The model consists of four layers and allows for limb deformation. With this model, our objective is to recover its parameters (and thus the human body pose) from automatically extracted silhouettes. LDM recovery algorithm is first developed for manual silhouettes, in order to generate ground truth sequences for comparison and useful statistics regarding the LDM parameters. It is then extended for automatically extracted silhouettes. The proposed methodologies have been tested on 10005 frames from 285 gait sequences captured under various conditions and an average error rate of 7% is achieved for the lower limb joint angles of all the frames, showing great potential for model-based gait recognition.

1 Introduction

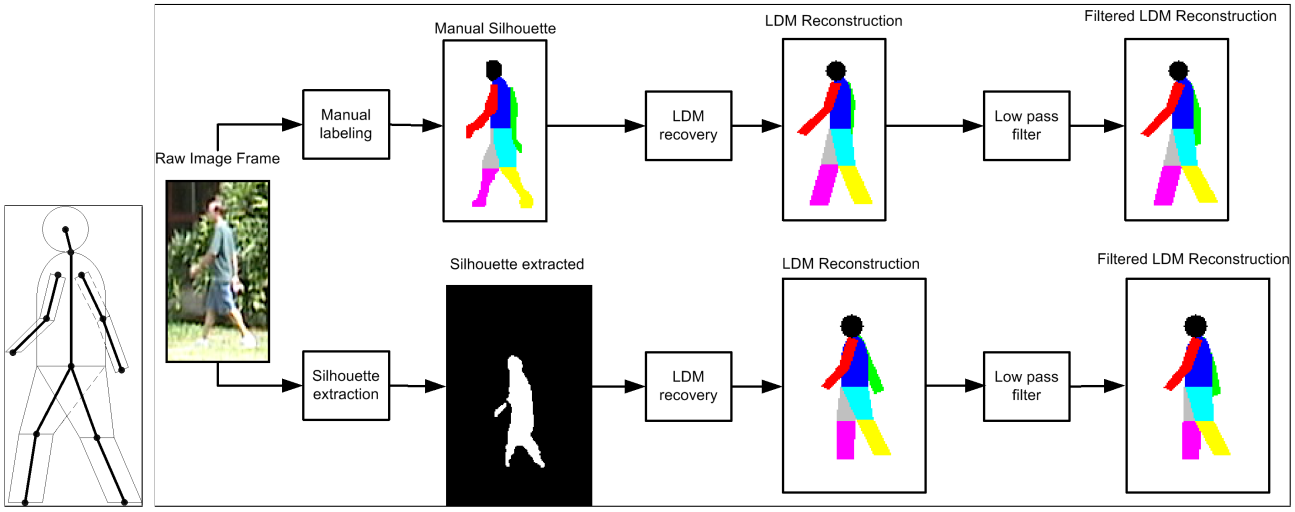
Gait recognition [9], the identification of individuals in video sequences by the way they walk, has recently gained significant attention. This interest is strongly motivated by the need for automated person identification system at a distance in visual surveillance and monitoring applications in security-sensitive environments, e.g., banks, parking lots, and airports, where other biometrics such as fingerprint, face or iris information are not available at high enough resolution for recognition [4]. Furthermore, night vision capability (an important component in automated surveillance at a distance) is usually impossible with other biometrics due to the limited signature in the IR image [4]. Gait, the particular way one walks, is a complex spatio-temporal biometric characteristic that can address the problems above. In contrast with those physiological biometrics such as face and iris, it is a behavioral (habitual) biometric with the following characteristics: uniqueness, unobtrusiveness and recognition at a distance (in low resolution video) [13].

In [10], Sakar *et al.* introduced the HumanID Gait Challenge problem, providing a set of twelve experiments of increasing difficulty, which examine the impact of five covariates on performance. The challenge provided the means to measure progress in the area and various researchers have reported results on these data sets [1, 5, 10]. While high recognition rates have been achieved on the easier sets, the recognition rates for the more difficult ones remain low. Articulated human body models are popular in human pose recovery and tracking [11, 16]. Gait recognition algorithms using 2D fronto-parallel body models are also proposed on some other data sets [3, 12–14], but to our knowledge, there are no studies and no results reported on the Gait Challenge data sets using articulated human body model.

In this paper, a new articulated human body model is proposed for gait analysis and model-based gait recognition. This model is called the layered deformable model (LDM) and it is inspired by the manually labeled silhouettes in [6]. The model is introduced in Sec. 2. In Sec. 3, algorithms that can be used to estimate the human body poses from the manual silhouettes to obtain the "ground truth" are introduced and analyzed. LDM recovery algorithm for automatically extracted silhouettes is also discussed in this section. The recovered parameters from both processes are smoothed by applying a number of constraints and a moving average filter, and they are compared to evaluate the performance, as shown in Fig. 1(b). The experimental results from processing 10005 frames using 285 Gait Challenge [10] sequences are reported in Sec. 4. Finally, conclusions are drawn in Sec. 5.

2 The Layered Deformable Model

In model-based gait recognition, the desirable model should be of moderate complexity allowing for fast process-



(a) LDM. (b) LDM human body pose recovery from manual and automatically extracted silhouettes.

Figure 1. The layered deformable model (LDM) and LDM pose recovery.

ing while at the same time provides enough features for discriminant learning. In other words, a trade-off between the computational complexity (the efficiency) of the model and the descriptive representation of the human gait (the accuracy) is sought. The desirable model is not to be as detailed as a fully deformable limb model while it must model limbs individually to enable model-based recognition.

This work introduces a human body model called the layered deformable model (LDM). While this model can be designed for gait from various viewing angles, the most commonly used fronto-parallel gait (side-view) is the main focus in this paper. Without loss of generality, it is assumed that the walking direction is from right to left. This model is inspired by the manually labeled silhouettes in [6].

The proposed LDM is illustrated in Fig. 1(a) and its close match to the manual silhouettes can be observed in Fig. 1(b). It consists of ten segments modeling ten body parts: head (circle), torso (semi-ellipse on top of a rectangle), left/right upper arms (rectangles), left/right lower arms (quadrangles), left/right upper/lower legs (quadrangles). Feet and hands are not modeled explicitly because in gait-based recognition tasks, they are relatively small in size and difficult to detect consistently due to occlusion. The model is defined based on a simple skeleton model. The skeleton is shown as thick lines and black dots in the figure.

The LDM is specified using 22 parameters that define the lengths, widths, positions and orientations of body parts:

- Lengths: the lengths of various body parts l_H (the radius of head), l_T (torso), l_{UA} (upper arm), l_{LA} (lower arm, including hand), l_{Th} (thigh) and l_{LL} (lower leg, including feet).
- Widths: the widths (thickness) of body parts w_T (torso, which is equal to the width of the top of thigh), w_K (knee) and w_A (arm, assuming the same width for upper and lower parts).
- Positions: the global position (x_G, y_G), which is also the position of the hip joint, and the shoulder displacement (dx_{Sh}, dy_{Sh}).
- Joint angles (body part orientations): θ_{lTh} (left thigh), θ_{rTh} (right thigh), θ_{lLL} (left lower leg), θ_{rLL} (right lower leg), θ_{lUA} (left upper arm), θ_{rUA} (right upper arm), θ_{lLA} (left lower arm), θ_{rLA} (right lower arm), and θ_H (head, neck joint angle).

For fronto-parallel gait, the lengths and widths of body parts do not vary much. Thus, the length and width parameters are considered static parameters that are consistent for the subject in a fronto-parallel gait sequence. The global position, shoulder displacement parameters and joint angles are dynamic parameters evolving with time, resulting in 13 degrees of freedom.

Furthermore, to model human body self-occlusion (e.g., between legs, arms and torso), the following four layers are introduced: right arm (layer one); right leg (layer two); head, torso and left leg (layer three); and left arm (layer four). Layer four is the closest to the camera (seldom occluded) and layer one is furthest from the camera (frequently occluded). Each layer and the resulted image are shown in Fig. 2 and self-occlusion is explained well with this model.

In addition, the LDM allows for limb deformation and Fig. 3 depicts an example of leg deformation. This is a significant difference from traditional 2D (rectangular) models and visual comparison with the manual silhouettes shows

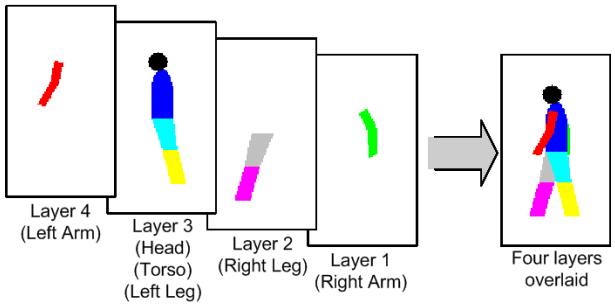


Figure 2. The layered representation.

that the LDM matches well with human's subjective perception of human body (in 2D).

Figure 3. Illustration of the deformation.

In summary, it can be argued that the LDM has a compact representation comparable to the simple rectangle/cylinder model [13] and its layered structure models self-occlusion between body parts. It allows for limb deformation while being considerably simpler than the deformable model of [16]. Moreover, the shoulder displacement parameters model shoulder swing observed in the manual silhouettes, and they also relate to viewing angles. On the whole, the LDM is able to model human gait realistically with moderate complexity and it resembles the manual silhouettes well.

3 LDM human body pose recovery

With the LDM, human body poses (LDM parameters) can be recovered for gait analysis. This problem is solved in two phases. LDM parameter estimation from the manual silhouettes is solved first. The obtained results serve as the ground truth for evaluating the automatic recovery algorithm. In addition, statistics from these ground truth data are used in the following task of obtaining the LDM estimations directly from the silhouettes extracted from raw gait sequences automatically.

3.1 Pose estimation from manual silhouettes

The manual silhouettes are available for a single gait cycle in each gait sequence labeled [6]. LDM parameters for a manual silhouette are first estimated by processing each individual body segment one by one. Some parameters, such as the limb joint angles, are more closely related to the way one walks and hence they are more important to gait recognition than the others, such as the width parameters. Thus, limb joint angle parameters are estimated first by employing robust algorithms to achieve high accuracy.

3.1.1 Estimation of limb joint angles

To reliably extract the limb joint angles (θ_{lTh}, θ_{rTh}, θ_{lLL}, θ_{rLL} , θ_{lUA}, θ_{rUA}, θ_{lLA} and θ_{rLA}), we propose to estimate the joint angles and positions of the limbs from reliable edge orientation, where they are estimated from either front or back edges only, decided by the current stance (posture). E.g., the front (back) edges are more reliable when the limbs are in front (at back) of the torso. This method of estimation through reliable body part information extends the leading edge method originally introduced in [14] so that noise and impairments due to loose cloths are greatly reduced. The mean-shift algorithm [2], a kernel-based algorithm for nonparametric mode-seeking, is applied in the joint spatial-orientation domain, taking care of different scales in the two domains by using different kernel sizes for different domains. This algorithm is applied to the reliable edges of each limb individually, preprocessed by a Gaussian filter to reduce noise. Let an edge pixel feature vector $\mathbf{p}_i = [\mathbf{p}_i^s; \mathbf{p}_i^o]$, where $\mathbf{p}_i^s$ is the spatial coordinate and $\mathbf{p}_i^o$ is the local orientation. Denote by $\{\mathbf{p}_i\}_{i=1:R}$ the R reliable edge pixel feature vectors. Their modes $\{\mathbf{q}_{i,c}\}_{i=1:R}$ (defined similarly) are sought by iteratively computing

$$\mathbf{q}_{i,j+1} = \frac{\sum_{i=1}^{R} \mathbf{p}_i \cdot k\left(\left\|\frac{\mathbf{q}_{i,j}^s - \mathbf{p}_i^s}{h_s}\right\|^2\right) \cdot k\left(\left\|\frac{\mathbf{q}_{i,j}^o - \mathbf{p}_i^o}{h_o}\right\|^2\right)}{\sum_{i=1}^{R} k\left(\left\|\frac{\mathbf{q}_{i,j}^s - \mathbf{p}_i^s}{h_s}\right\|^2\right) \cdot k\left(\left\|\frac{\mathbf{q}_{i,j}^o - \mathbf{p}_i^o}{h_o}\right\|^2\right)} \quad (1)$$

until convergence, where $k(\cdot)$ is a kernel, and h_s and h_o are the kernel bandwidths for the spatial and orientation domain, respectively, with the initialization $\mathbf{q}_{i,1} = \mathbf{p}_i$. The modes (points of convergence) are sorted in descending order based on the number of points converged to each mode. The dominant modes (modes at the top of the list) represent body part orientations and insignificant modes (modes at the bottom of the list) are ignored.

3.1.2 Estimation of other parameters

With the limb joint angles estimated, the joint (e.g., elbow, knee) positions can be determined easily and the lengths (l_{UA}, l_{LA}, l_{Th} and l_{LL}) and widths (w_K and w_A) are estimated from them. From the bounding box of the torso segment, w_T, l_T and (x_G, y_G) are estimated. For the head, the "head top" and "front face" points are estimated through Gaussian filtering and averaging, and they determine the head size (l_H) and the head center, partly eliminating the effects of hair styles. θ_H can then be estimated from the

head center and the neck joint position (estimated from torso), and (dx_{Sh}, dy_{Sh}) is determined from the difference between the neck and shoulder joint positions.

3.1.3 Post-processing of the estimation results

Due to the imperfection of manual labeling and the pose recovery algorithm above, the estimated LDM parameters are not varying smoothly and they need to be smoothed within a gait sequence. Thus, a two-step post-processing is proposed. The first step applies a number of constraints. Inter-frame parameter variation limits are applied and the head size (l_H) is fixed to be the median over a cycle. The inter-dependence between angles of the same limbs are enforced to realistic values by respecting the following conditions:

$$\theta_{lTh} \leq \theta_{lLL},\ \theta_{rTh} \leq \theta_{rLL},\ \theta_{lUA} \geq \theta_{lLA},\ \theta_{rUA} \geq \theta_{rLA}. \quad (2)$$

In the second step, a moving average filter of window size $n = 2k + 1$ is applied to the parameter sequences by replacing the estimation of a parameter z_t with: $\overline{z_t} = \frac{1}{n}\sum_{i=-k}^{k} z_{t+i}$.

3.2 Automatic pose estimation

Automatic pose recovery from a given gait sequence is discussed in this section. Several background subtraction algorithms are available to obtain silhouettes automatically from a video sequence. A coarse-to-fine silhouette extraction algorithm is developed in this work based on the work proposed in [8], and it is described in [7].

3.2.1 Estimation of static parameters

Since static parameters are largely affected by cloths and the silhouette extraction algorithm used, they are not considered as features useful for model-based gait recognition and coarse estimations are used for them. The statistics of their ratios to the silhouette height h_S are studied for the Gallery set of manual silhouettes and the standard deviations in these values are found to be low. Therefore, fixed ratios to h_S are used for the static parameter estimations, based on the Gallery set of manual silhouettes.

3.2.2 Automatic silhouette information extraction

With help from the ideal proportions of the human (eight-head-high) figure in drawing [15], the following information is extracted for LDM parameter (pose) estimation from the automatically extracted silhouettes: the silhouette height h_S; the first row y_{min} and the last row y_{max} of the silhouette; the center column c_H of the first $h_S/8$ rows (for head position); the center column of the waist c_W, which is the average column position of the rows $h_S/8$ to $h_S/2$ (the torso portion) with widths within ± 0.3 deviation from the expected torso width $0.169 \cdot h_S$; the joint spatial-orientation domain modes and the number of points converged to each mode for the front and back edges of limbs, which are obtained using mean-shift as in Sec. 3.1.3. For the upper arms (rows $h_S/8$ to $3 \cdot h_S/8$), due to significant collusion with the torso, only the front edge of the left upper arm and the back edge of the right upper arm are used.

3.2.3 Estimation of dynamic parameters

Firstly, the global position (x_G, y_G) is determined as

$$x_G = c_W,\ \ y_G = y_{min} + 2 \cdot l_H + l_T. \quad (3)$$

θ_H is then calculated through estimating the neck joint $(x_G, y_G - l_T)$ and the head centroid $(c_H, y_{min} + l_H)$. Next, the joint angles are estimated. The left/right limb angles in this section refer to the angles estimated for the left/right limb in the silhouettes, respectively. The angles θ_{lLL}, θ_{rLL}, θ_{lLA} and θ_{rLA} are estimated by examining the difference of the front and back edge estimations and the variation of the estimations (compared to those in the last frame). The knee positions, with row number set to $(y_{max} - h_S/4)$, are determined using the mean-shift modes of the lower legs. Then, θ_{lTh} and θ_{rTh} are calculated from the hip (x_G, y_G) and knee joint positions, and θ_{lUA} and θ_{rUA} are set to the estimations from Sec. 3.2.2. The left elbow and shoulder positions are obtained similarly from the left arm and (dx_{Sh}, dy_{Sh}) is determined. The constraints described in the first step of post-processing in Sec. 3.1.3 are enforced in this estimation process and a number of other heuristics/constraints are applied to improve the results.

3.2.4 Limb switching detection

The correct labeling of left/right limbs needs to be addressed for accurate pose recovery. Since most gait recognition algorithms do not differentiate left and right legs, without loss of generality, it is assumed that in the first frame, the left and right legs are "switched", i.e., the left leg is on the right and the right leg is on the left. Next, we attempt to label the limbs correctly in subsequent frames. When the thighs switch and when the lower legs switch are determined by examining the variations of respective joint angles. The arms switch in opposite direction of the thighs since in normal gait, the arms have the opposite "switching" mode. In addition, we set the minimum number (11) of frames between two successive switches.

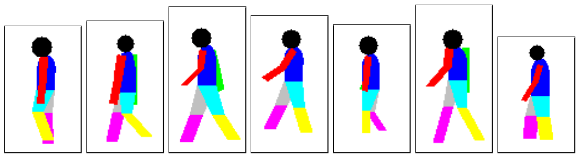
Finally, the estimated values are smoothed through the two-step post-processing proposed in Sec. 3.1.3. The performance of the proposed LDM recovery algorithms is presented in the next section in details.

4 Experimental results

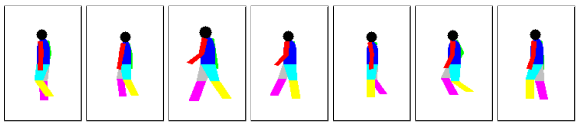
The proposed solutions have been tested on all the manual silhouettes in [6]. There are 285 sequences from five data sets (Gallery and Probes B, D, H, and K) and each sequence consists of one gait cycle (around 36 frames). The automatic pose estimation algorithm is applied to the corresponding silhouettes automatically extracted. For the mean-shift, we set $h_s = 15$ and $h_o = 10$ and use the kernel with Epanechnikov profile [2]. In this case, the mean-shift reduces to simple (weighted) average. For the running average filter, a window size $n = 7$ is used.

Several examples are shown in Fig. 4 to assist in the qualitative evaluation of the performance. The first five examples are successful examples taken from each of the five sets considered here. The last two are examples of poor estimation results. The large estimation error in the sixth example is due to silhouette extraction noise, while the incorrect labeling of the legs in the last example can be attributed to both silhouette noise and the limb switching algorithm's failure. We are going to investigate algorithms to reduce such errors in the future, e.g., by providing feedback from the LDM recovery to the background subtraction process.

(a) The raw image frames.



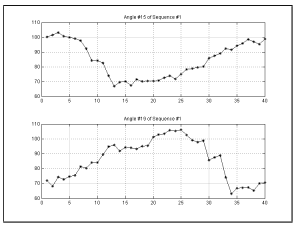
(b) The LDM recovery for manual silhouettes.



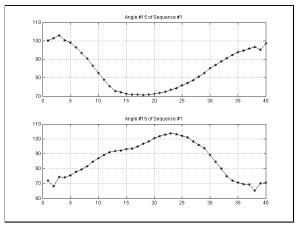
(c) The LDM recovery for automatically extracted silhouettes.

Figure 4. Examples of the recovered poses.

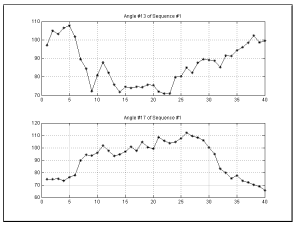
Figure 5 depicts the improvement in joint angle estimation, which can be obtained through the utilization of a simple average filter. One gait sequence is shown here with the top and bottom subplot in each sub-figure showing the left and right thigh joint angles recovered, respectively. The simple filter appears to provide good smoothing results. Moreover, similarity between the filtered angles can be observed, which is promising to be used for gait recognition in our future work.



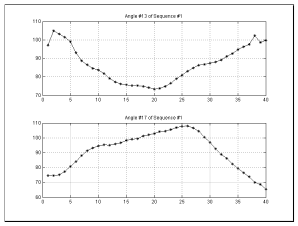
(a) Angles from the manual silhouettes before filtering.



(b) Angles from the manual silhouettes after filtering.



(c) Angles from the extracted silhouettes before filtering.



(d) Angles from the extracted silhouettes after filtering.

Figure 5. Joint angle filtering.

In terms of quantitative analysis, the error rate (in percentage) of joint angle estimations is calculated as following:

$$l(\theta_s, \theta_m) = 100 \cdot |\theta_s - \theta_m| \,/\, |\theta_m| \,, \qquad (4)$$

where θ_s represents the angle recovered from the extracted silhouettes and θ_m represents the angle recovered from the manual silhouettes. Table 1 gives the average error rate for the joint angle estimations. We are particularly interested in the estimated lower limb joint angles since they are important for model-based gait recognition. There are six sub-figures shown in Fig. 6. The first five sub-figures show the error rates for the Gallery and Probes B, D, H and K sets. The last sub-figure shows the average over these five sets. The x axis represents the left thigh (1), the left lower leg (2), the right thigh (3) and the right lower leg (4) angles. An average error rate of 7% is achieved for the lower limb joint angles and the fifth x value in the last sub-figure shows this. This encourages us to explore the potential of our algorithms in model-based gait recognition.

5 Conclusions

Recently, gait recognition has attracted much attention for its potential in surveillance and security applications. The release of the Gait Challenge data sets provides a common database for testing and evaluation of gait recognition algorithms. This paper proposes a layered deformable model for gait analysis, with 22 parameters defining the body part lengths, widths, positions and joint angles of human body. Algorithms are proposed to recover human body

Table 1. The average error rate for the joint angle estimations.

Angles	θ_{lTh}	θ_{lLL}	θ_{lUA}	θ_{lLA}	θ_{rTh}	θ_{rLL}	θ_{rUA}	θ_{rLA}	θ_H
$l(\theta_s, \theta_m)$ (%)	7.5237	6.8135	19.125	17.918	7.3876	7.0189	14.26	14.534	10.422

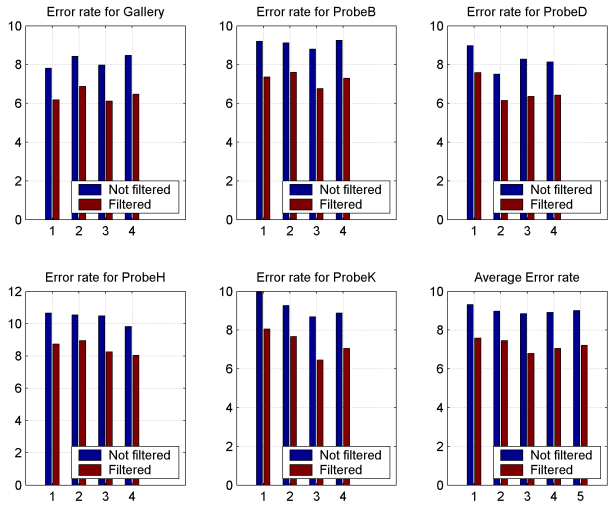


Figure 6. The error rate of model recovery.

poses from the manual silhouettes (the ground truth) and the automatically extracted silhouettes, and the estimations from these two sets are compared to evaluate the performance. The algorithms work well on the whole though with some exceptions. Experiments show that an average error rate of 7% is achieved for the lower limb joint angles, which are important for model-based gait recognition. Naturally, we will continue to improve on the estimation algorithms and work on gait recognition using these estimated angles in our future.

Acknowledgment

The authors would like to thank Prof. Sarkar from the University of South Florida (USF) for providing the manual silhouettes and the Gait Challenge data sets. Support provided by the Communications and Information Technology Ontario Partnership Program and the Bell University Labs - at the University of Toronto is also acknowledged.

References

[1] N. V. Boulgouris, D. Hatzinakos, and K. N. Plataniotis. Gait recognition: a challenging signal processing technology for biometrics. *IEEE Signal Processing Mag.*, 22(6):78–90, Nov. 2005.

[2] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Machine Intell.*, 24(5):603–619, May 2002.

[3] D. Cunado, M. S. Nixon, and J. N. Carter. Automatic extraction and description of human gait models for recognition purposes. *Computer Vision and Image Understanding*, 90(1):1–41, Jan. 2003.

[4] A. Kale, A. N. Rajagopalan, A. Sunderesan, N. Cuntoor, A. Roy-Chowdhury, V. Krueger, and R. Chellappa. Identification of humans using gait. *IEEE Trans. Image Processing*, 13(9):1163–1173, Sept. 2004.

[5] L. Lee, G. Dalley, and K. Tieu. Learning pedestrian models for silhouette refinement. In *Proc. IEEE Conf. on Computer Vision*, pages 663–670, Oct. 2003.

[6] Z. Liu, L. Malave, and S. Sarkar. Studies on silhouette quality and gait recognition. In *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, volume 2, pages 704–711, 27 June-2 July 2004.

[7] H. Lu, K. Plataniotis, and A. Venetsanopoulos. Coarse-to-fine pedestrian localization and silhouette extraction for the gait challenge data sets. In *Proc. IEEE Int. Conf. on Multimedia & Expo*, 2006. submitted.

[8] J. Migdal and W. E. L. Grimson. Background subtraction using markov thresholds. In *IEEE Workshop on Motion and Video Computing*, Jan. 2005.

[9] M. S. Nixon and J. N. Carter. Advances in automatic gait recognition. In *Proc. IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pages 139–144, May 2004.

[10] S. Sarkar, P. J. Phillips, Z. Liu, I. Robledo, P. Grother, and K. W. Bowyer. The human ID gait challenge problem: Data sets, performance, and analysis. *IEEE Trans. Pattern Anal. Machine Intell.*, 27(2):162–177, Feb. 2005.

[11] H. Sidenbladh, M. J. Black, and D. J. Fleet. Stochastic tracking of 3d human figures using 2d image motion. In *Proc. European Conference on Computer Vision*, pages 702–718, June 2000.

[12] D. K. Wagg and M. S. Nixon. On automated model-based extraction and analysis of gait. In *Proc. IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pages 11–16, May 2004.

[13] L. Wang, H. Ning, T. Tan, and W. Hu. Fusion of static and dynamic body biometrics for gait recognition. *IEEE Trans. Circuits Syst. Video Technol.*, 14(2):149–158, Feb. 2004.

[14] C. Y. Yam, M. S. Nixon, and J. N. Carter. Automated person recognition by walking and running via model-based approaches. *Pattern Recognition*, 37(5):1057–1072, May 2004.

[15] A. Zaidenberg. *Drawing the figure from top to toe*. World publishing company, 1966.

[16] J. Zhang, R. Collins, and Y. Liu. Representation and matching of articulated shapes. In *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pages 342–349, 27 June-2 July 2004.