Human Gait Recognition

[1]Rong Zhang [2]Christian Vogler [1]Dimitris Metaxas

[1] Department of Computer Science
Rutgers University
110 Frelinghuysen Road
Piscataway, NJ 08854

[2] Gallaudet Research Institute
Gallaudet University
HMB S-433, 800 Florida Ave. NE
Washington, DC 20002

Abstract

The reliable extraction of characteristic gait features from image sequences and their recognition are two important issues in gait recognition. In this paper, we propose a novel 2-step, model-based approach to gait recognition by employing a 5-link biped locomotion human model. We first extract the gait features from image sequences using the Metropolis-Hasting method. Hidden Markov Models are then trained based on the frequencies of these feature trajectories, from which recognition is performed. As it is entirely based on human gait, our approach is robust to different type of clothes the subjects wear. The model-based gait feature extraction step is insensitive to noise, cluttered background or even moving background. Furthermore, this approach also minimizes the size of the data required for recognition compared to model-free algorithms. We applied our method to both the USF Gait Challenge data-set and CMU MoBo data-set, and achieved recognition rate of 61% and 96%, respectively. The results suggest that the recognition rate is significantly limited by the distance of the subject to the camera.

1. Introduction

Human recognition is an important task in a variety of applications, such as access control, surveillance, etc. To distinguish different persons by the manner they walk is a natural task people perform everyday. Psychological studies [10, 19] have showed that gait signatures obtained from video can be used as a reliable cue to identify individuals. These findings inspired researchers in computer vision to extract potential gait signatures from images to identify people. It is challenging, however, to find idiosyncratic gait features in marker-less motion sequences, where the use of markers is avoided because it is intrusive and not suitable in general gait recognition settings.

Ideally, the recognition features extracted from images should be invariant to factors other than gait, such as color, texture, or type of clothing. In most gait recognition approaches [6, 16, 11], recognition features are extracted from silhouette images. Although these features are invariant to texture and color, the static human shape, which is easy to be concealed, inevitably mingles with the movement features. In this paper, we propose a 2-step, model-based approach, in which reliable gait features are extracted by fitting a five-link biped human locomotion model for each image to avoid shape information, followed by recognition using Hidden Markov Models (HMMs) based on the frequency components of the trajectories of the relative joint positions. Applying our approach to both the USF Gait Challenge data-set and the CMU MoBo data-set, we demonstrate that promising recognition rate can be obtained using gait only features.

This paper is organized as follows. Section 2 summarizes the existing approaches to the gait recognition problem. The five-link biped human model is described in Section 3. Section 4 provides details of the extraction of gait features, while recognition using HMM is described in Section 5. Experimental results are presented in Section 6, followed by conclusions in Section 7.

2. Previous Approaches to Gait Recognition

Existing methods for gait recognition can be divided in two main categories: model-free and model-based.

Two model-free baseline approaches have been proposed for gait recognition based on the silhouette images: Phillips et al. [16] measured the correlation between the probe silhouette image sequences and those in a data-set, while Collins et al. [3] applied template matching between selected key frames. Other low level image features are extracted, for identification of the spatial and temporal variances of human gait: the width of the outer contour of the silhouette [11], gait mask responses [6], moments of the optical flows [14], generalized symmetry operator [8], etc. Lee et al. [13] fit seven ellipses in the human body area, and used their locations, orientations, and aspect ratios as

Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW'04)
1063-6919/04 $ 20.00 IEEE

features to represent the gait. All features used in these approaches are calculated either on the pixel level (background subtraction) or within small regions (edge map and optical flow calculation), hence are susceptible to noise and background cluttering. More recently, silhouette refinement [12] has been proposed to improve the recognition rate. However, the features extracted using the above methods include shape information, which should be avoided for gait recognition.

On the other hand, with gait features closely related to the walking mechanics, model-based approaches have the potential for robust feature extraction. Human-like structures are proposed in gait feature extraction: A 2D-stick model was obtained through line fitting to the skeleton of the silhouette images by Niyogi and Adelson [15], while Cunado et al. [4] modelled the thighs as interlinked pendular to extract their angular movements. These recognition features are extracted over large regions, which are less sensitive to the image noise. More significantly, theses features do not contain shape information. The above methods achieved satisfactory recognition rate over small gait data set, demonstrating the potential of identifying persons using movement features only.

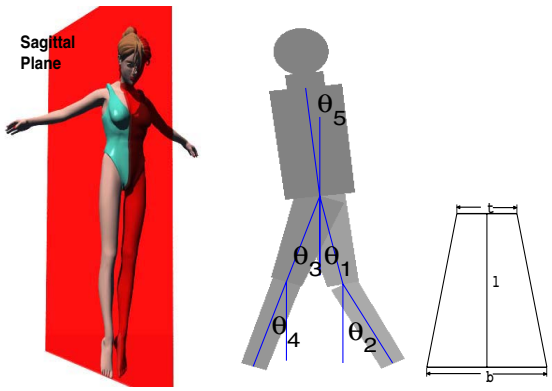
3. Five-link Biped Model

Walking is a complex dynamic activity. A good human model for gait recognition should be simple, but general enough to capture the dynamics of most pedestrians, and to be customized to fit different persons in tracking sequences. Complicated human models, such as the 3D deformable model [18], are not practical for efficient human tracking.

Studies carried out by physiologists show [2] that most walking dynamics take place in the sagittal plane, or the plane bisecting the human body [Figure 1 (a)]. Hence, we use a 2D five-link biped locomotion model to represent the human body in the image sequences when the person is walking parallel to the camera (side view).

The biped model is constructed as shown in Figure 1. The lower limbs are represented as trapezoids, whereas the upper body is simplified as the upper half of the human silhouette without arms. For people walking at a distance, it is difficult to recover the exact arm positions, as little information of the arms is available in the visual images. Therefore, the influence of the arm dynamics has been neglected in our dynamic model. These simplifications are necessary for a compact model to reduce the computational complexity, while at the same time enabling capturing most dynamics of pedestrians.

If the length of each part (shape model) is fixed, the biped model M has seven degrees of freedom, $M = (C = \{x, y\}, \Theta = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\})$, where C is the position of the body center in the image, and Θ is the orientation vector consisting of sagittal plane elevation angles (SEAs) for the five body parts. The SEAs of a certain body part is defined as the angle between the main axis of the body part and the y axis [20], as shown in Figure 1 (b). However, the size of each body part in the images may differ with different persons, or even the same person at different distances from the camera. One way to obtain the shape model for each person is to manually find the joint positions on the first image, which is tedious when the data set is huge. In our work, we develop a model fitting method for the initialization process, in which the size of each body part in the image is specified by fitting the human shape model to the silhouette image.



(a) (b) (c)

Figure 1: (a) Sagittal plane. (b) Five-link biped human model. (c) An individual body part

3.1. Scale-invariant body model

For our purpose, we need a general human shape model independent of scaling. As shown in Figure 1, the human body model, without considering neck and head, consists of five trapezoids, connected at the joints. Each trapezoid is defined by its height (l) and the lengths of the top and bottom bases (t and b, respectively). Hence, each body part p_i, $i = 1, \ldots, 5$, can be represented by $p_i = \{\alpha_i, \beta_i, l_i\}$, where $\alpha = t/l$ and $\beta = b/l$ are base-to-height ratios. By normalizing the body part heights with respect to the height of the trunk (l_5), we obtain a shape model invariant to scaling, which is parameterized by two vectors: the base-to-height ratio vector, $K = \{\alpha_1, \beta_1, \alpha_2, \beta_2, \ldots, \alpha_5, \beta_5\}$, and the relative height vector $R = \{r_1, r_2, \ldots, r_5\}$, where $r_i = l_i/l_5$. Together with the biped model M, we can describe the human body posture as $H = \{K, R, M\}$.

We assume that the model parameters are independent of each other, subject to Gaussian distributions. The orientation vector is subject to a uniform distribution over an interval L_Θ, provided by physical limits of the joints. Thus, the probabilistic distribution of the human model can be ex-

Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW'04)
1063-6919/04 $ 20.00 IEEE

pressed as

$$H = \{K, R, M\} \sim G(K, \Sigma_K)G(R, \Sigma_R)U(C)U(\Theta; L_\Theta) \quad (1)$$

The means and variances are estimated from the measurements provided in [21].

3.2. Initialization of body shape model

The orientation and the actual length of each body part in the image are defined in the initialization step. We choose to fit the silhouette image when the two legs are furthest apart from each other (double stance phase). In this phase, the SEAs of the shank and calf of the same leg are roughly identical.

We first obtain a rough estimation of the human model H from the silhouette image S. The body center position in the image is set to the middle point of the silhouette pixels

$$C = (x, y) = (\underset{x_i \in S}{median}(x_i), \underset{y_i \in S}{median}(y_i)). \quad (2)$$

To calculate the orientation vector Θ, we select three sub-regions within the silhouette image, one for the upper body and one for each leg as shown in Figure 2 (a). The SEA of each body part is set as the angle of the main axis for pixels within the corresponding region, which is the axis with the least second moment [9]. Given Θ, the height of each body part can be obtained based on the height of the silhouette image.

The above estimation provides a good approximation, however it may not be accurate in some cases. For further adjustment, we seek for a human model H^* which fits the silhouette image S best. Using Bayesian inference, we formulate this procedure as:

$$\begin{aligned} H^* &= arg\,\underset{H}{max}\, p(H|S) \\ &= arg\,\underset{H}{max}\, p(S|H)p(H) \end{aligned} \quad (3)$$

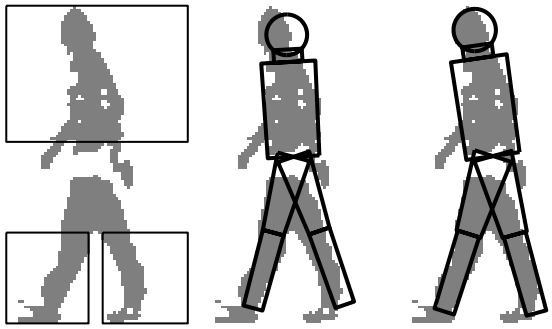
where the prior distribution $p(H)$ is given in Eq. (1), and the likelihood $p(S|H)$ specifies the silhouette generating process from human model H to S.

Assume S' is the shape generated by the human model H, C_S is the boundary point set for shape S, and $\mathcal{A}(\mathcal{S})$ is the corresponding area. The likelihood function is defined as

$$\begin{aligned} &p(S|H) = p(S|S') \\ &\sim (\prod_{v \in C'_S} G(D(v, C_S), \sigma_d^2))^{w_1} (G(\mathcal{A}(S) - \mathcal{A}(S'), \sigma_S^2))^{w_2} \end{aligned} \quad (4)$$

where σ_d^2, and σ_S^2 are the variances of the distance and area, respectively, w_1 and w_2 are the weights for the two components, and

$$D(v, C_s) = \min_{v' \in C_S} d(v, v') \quad (5)$$



(a) (b) (c)

Figure 2: (a) Silhouette image. The three blocks indicate regions for shape model calculation. (b) Rough estimation result. (c) Initialization result

is the minimum distance from point v to the contour of S, calculated by the distance transform.

Finding the global optimal H^* is rather difficult. Here we use the Metropolis-Hasting method, which guarantees convergence to sample from the posterior. Starting from the rough estimation H obtained above, the Metropolis-Hasting steps for adjusting H are:

1. Generate new sample H' according to $q(H \rightarrow H')$.

$$q(H \rightarrow H') \propto p(H')G(H - H', \Sigma_H), \quad (6)$$

   where Σ_H is the covariance matrix for model parameters.

2. Accept H' according to

$$\alpha = min(1, \frac{p(H'|S)q(H' \rightarrow H)}{p(H|S)q(H \rightarrow H')}).$$

3. Repeat step 1 and 2 until $p(H|S)$ is high enough or a maximum number of iterations is reached.

The sizes of neck and head are then calculated through the relative size with respect to the trunk length based on [21]. Figure 2 (c) shows the initialization result, which shows improvement from the rough estimation in Figure 2 (b). In addition, we can obtain the appearance model (W) of each body part based on the color information within the corresponding image region. The initialization step relies on the quality of the silhouette image; therefore, further adjustment of the parameters may be needed if the silhouette image is severely corrupted.

4. Tracking

Since we have extracted the shape model and the initial configuration, the next step is to extract gait signatures over

time based on this shape model. Current 2D-based tracking methods use either image edges or dense optical flow for detection and tracking. However, image cues, such as optical flows and edges, are not totally reliable, especially when calculated from noisy images. To achieve robustness, we need to carry out our computations within large regions, e.g., at the body part level. The image information we utilize is the color and the inner silhouette region.

For an input frame I_t at the time instance t, we use the background model to obtain the silhouette image S_t. Given the appearance model (W) and human model parameters $M_t = (C_t, \Theta_t)$, we can compose an image $I(M_t; W)$. The best human model configuration should make this image as close to I_t as possible. In addition, the area of the human model should be equal to the area of the silhouette image, and the difference of the biped model configurations between time instance $t-1$ and t is small. Therefore, we want to estimate the best biped model M_t which minimizes the total energy of the following equation,

$$\begin{aligned} E =& w_c \sum \rho(I_t - I(M_t; W), \sigma) \\ &+ w_{\mathcal{A}}(\mathcal{A}(S_t) - \mathcal{A}(M_t))^2 + w_m |M_t - M_{t-1}| \end{aligned} \tag{7}$$

where w_c, $w_{\mathcal{A}}$, and w_m are three weight factors, ρ is the Geman-McClure function defined as [1]:

$$\rho(x, \sigma) = \frac{x^2}{\sigma^2 + x^2}, \tag{8}$$

which is robust error norm since it constrains the effect of large residue (x) value. The scale parameter σ is defined as:

$$\sigma = 1.4826 \times median|I_t - I(M_t; W)| \tag{9}$$

The initial C_t is calculated as the mass center of the silhouette image as given in Eq. (2). The predicted orientation are given by:

$$\Theta_t = 2 * \Theta_{t-1} - \Theta_{t-2} \tag{10}$$

The minimization of the energy term in Eq. (7) is maximizing the following probability

$$p(M_t|I_t) \propto exp(-E), \tag{11}$$

by employing the same Metropolis-Hasting method used in the initialization step.

5. Recognition

The sagittal elevation angles extracted from the above tracking procedure capture the temporal dynamics of the gait of the subject, whereas the trajectories of the corresponding joint position reveal the spatial-temporal history. In addition, studies [2, 20] showed that the SEAs exhibit less inter-subject variation across humans. Therefore, our recognition method focuses on the joint position trajectories. In this section, we provide details of the recognition process.

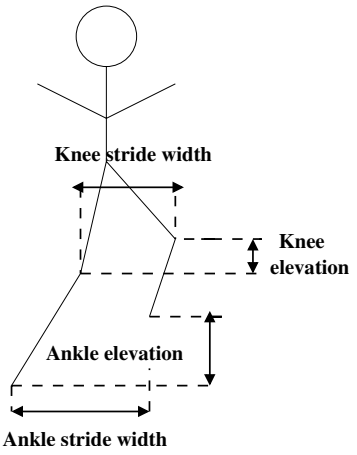


Figure 3: The space domain features

5.1. Recognition Features

Based on the tracking results obtained with the biped model, the differences across people are largely temporal. It is, therefore, necessary to choose a feature representation that makes the temporal characteristics of the data signal explicit. Because gait is cyclic, a frequency domain-based representation seems particularly suitable.

To this end, we first compute the following space domain features: ankle elevation (s_1), knee elevation (s_2), ankle stride width (s_3), and knee stride width (s_4) (Figure 3). The trajectories of the above four features within normalized gait cycle are shown in Figure 4.

For each of these four features s_i, we compute the Discrete Fourier Transform (DFT), denoted as S_i over a fixed window size of 32 frames which we slide over the feature signal sequences.

$$S_i(n) = \frac{1}{32} \sum_{k=0}^{31} s_i(k) e^{-2\pi i n k/32} \; for \; n = 0, \cdots, 31 \tag{12}$$

The size of 32 frames was chosen close to a typical human gait cycle. Future work should also investigate adaptive window size based on the actual period of the gait cycle of different person.

The DFTs reveal periodicities in the feature data as well as the relative strengths of any periodic components. Since that the lowest frequency component does not provide any information on the periodicity of the signal, while high frequency components mainly capture the noise, we sample the magnitude and phase of the second to fifth lowest frequency components. This leads to a feature vector containing 4 magnitude and 4 phase measures for each of the four space domain base features ($S_1, \cdots, S_4$), an overall dimension of 32.

Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW'04)
1063-6919/04 $ 20.00 IEEE

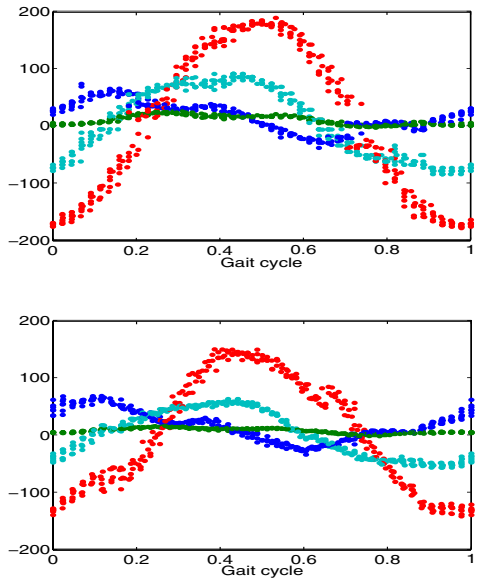


Figure 4: The trajectories of space domain features for two different subjects

5.2. Recognition Method

After computing the features, for each of the gait data samples, we then segment the resulting data stream according to its gait cycles, such that any single example contains only the data from a single gait cycle. In this way, recognition is analogous to isolated speech recognition. Therefore, we applied the Hidden Markov model (HMM) for identification, which have been used successfully in speech recognition.

We consider HMM of degree one, where the current state depends only on the previous state. The observation for HMM is the 32-dimensional feature described above. The HMM is represented as (π, A, B). In the training step, the initial state distribution π, transition probabilities A and the observation probability B are estimated using the standard Baum-Welch re-estimation method [17].

Given a test example, we compute the likelihood of each HMM on the example, and choose the HMM with the highest likelihood as the correct one, i.e., label it as k^* such that $k^* = \arg\max_k p_k(O|\pi, A, B)$.

With multiple gait cycles for the same person, we can recognize each gait cycle individually using above method. To combine the recognition results together, we aggregate the N-best recognition: for the recognition result of each gait cycle, assign a score of 20 to the first rank, 19 to the second rank, and so on; and then sum up all the rank scores for all possible hypotheses, and pick the result with the highest cumulative score. Performing aggregation in this way yields an improved identification rate.

(a) (b)

Figure 5: (a) Sample image from CMU MoBo data-set. (b) Result for fitting biped model

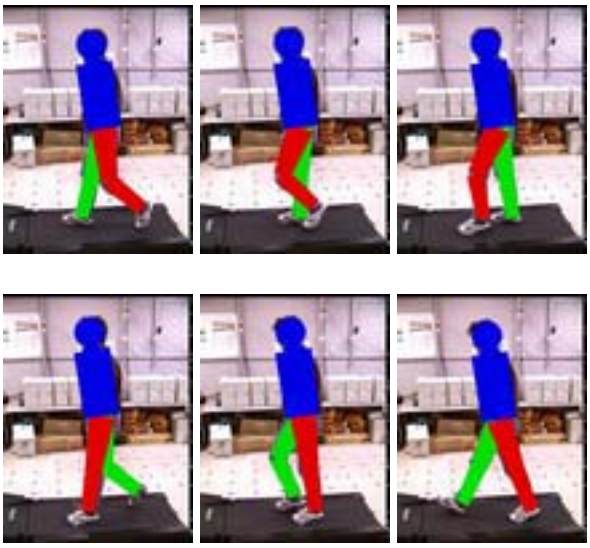


Figure 6: Tracking results for one subject

6. Experiments and results

The above algorithm is applied to both the CMU MoBo data-set [7] and the USF Gait Challenge data set [16].

6.1. CMU MoBo data-set

The CMU data-set contains video of 25 individuals with 824 gait cycles, who are walking on a treadmill under four different conditions: slow walk, fast walk, incline walk and walking with a ball in hand. Figure 5 and Figure 6 shows the tracking result of sample images.

In our first experiment, we split the gait cycles randomly into a training and a test set by a ratio of 3:1, so that both sets contain a mix of examples from all four walking activities. The results are shown in Figure 7. We achieve 96%

Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW'04)
1063-6919/04 $ 20.00 IEEE

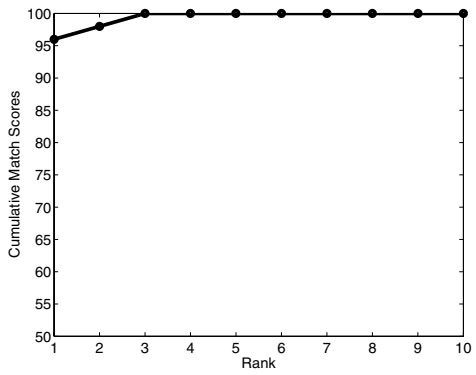
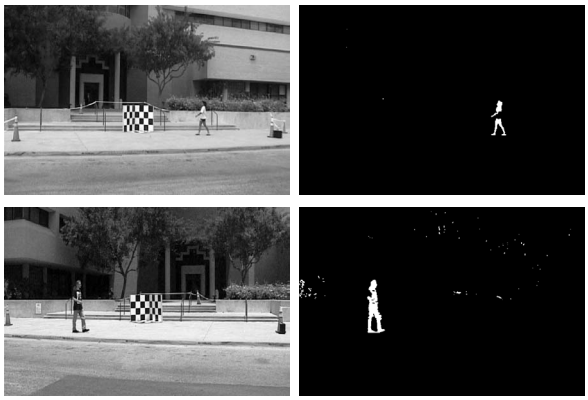


Figure 7: CMS plot of CMU gait data



(a) (b)

Figure 8: (a) Original images (b) Silhouette images

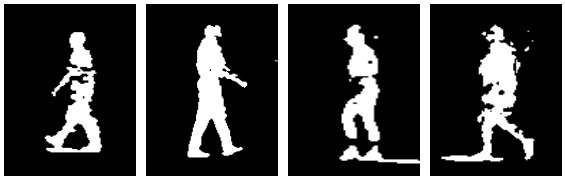


Figure 9: Sample silhouettes from the USF data

identification accuracy (Rank = 1), and the correct identification always occurs within top 3 ranks.

We also carried out the following experiments on this data-set:

1. Train with slow walk and test with slow walk.
2. Train with fast walk and test with fast walk.
3. Train with incline walk and test with incline walk.
4. Train with walking while holding a ball and test with walking while holding a ball.
5. Train with slow walk and test with walking while holding a ball.

In the first four experiments, the sequences are divided into two training and testing sets, by the ratio of 4:1. In case (5), the entire slow walk sequences are used for training, and only one gait cycle for walking with a ball is used for evaluation.

The results of the five experiments are shown in Table 1. As we can see, the recognition rate hit 100% at the top match for experiments (1) and (4), and is nearly perfect (around 96%) for (2) and (3). This shows our method even perform better than those shape-based approach such as [11], which suggests the motion dynamics of different subject are accurately captured using our method. However, our recognition result is much worse for experiment (5) than (1)−(4). Observed that while holding a ball, the subject slightly changes his/her gait due to the necessary adjustment to balance the additional weight. As a result, poor recognition rate is a natural outcome for methods using only movement information. This also indicates that the higher recognition rates achieved using other recognition methods employ the human shape information, in addition to gait.

Train vs Test	$P_I(\%)$ (at rank)			
	1	2	5	10
Slow vs Slow	100	100	100	100
Fast vs Fast	96.0	100	100	100
Incline vs Incline	95.8	100	100	100
Ball vs Ball	100	100	100	100
Slow vs Ball	52.2	60.9	69.6	91.3

Table 1: Performances for CMU MoBo data-set in terms of the identification rate P_I at ranks 1, 2, 5 and 10

6.2. USF Gait Challenge data-set

For the USF Gait Challenge data-set, since they are taken outdoors, we need to handle shadow, moving background, lighting changes, etc., in the silhouette extraction step. Therefore, we applied the non-parametric background modelling [5] for silhouette extraction. Figure 8 and Figure 9 show the silhouette images.

Typically, each gait data sample in this data-set contains 4 to 7 individual gait cycles. Overall, there are 75 subjects in the data set, with a total of 2045 gait cycles. 75% of the cycles are randomly selected to form the training set, with the rest forming the test set. Both sets contain a mix of examples from the subjects with different camera views, types of shoes and surfaces. The identification rate for the entire USF data-set is 61%, shown in Figure 10.

The recognition rate is lower than that for the CMU data-set, which may be attributed to either the number of subjects

Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW'04)
1063-6919/04 $ 20.00 IEEE

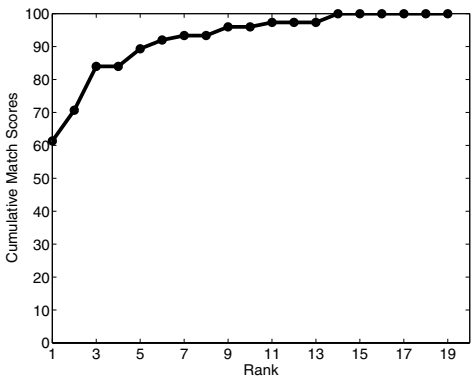


Figure 10: CMS plot of USF gait data

in the data-set, or the distance between the subjects and the camera. We randomly pick 25 subjects from the USF data-set and use their data for the recognition process. This experiment is carried out 20 times, and we obtained an average recognition rate of 77% with a standard deviation of 5%. Therefor, if we use only 25 subjects, the recognition rate is better than with 75 subjects, although this factor by itself does not fully account for the lower recognition rate for the USF data set. We notice that the average image length of thighs for the subjects is 26.7 pixels, while ~ 130 in CMU data-set. Therefore, the accuracy of the extracted feature is limited in USF data-set. The subtle inter-subject movement differences could not be fully extracted from these images, which results in lower recognition rate. Both the number of subjects and the image resolution are hence important in affecting the recognition rate.

7. Conclusion

In this paper, we have shown a novel 2-step, model-based approach for gait recognition using human body movements exclusively. As the concealable shape information is avoided, our method is more robust than the shape-based ones. Applying this approach to CMU MoBo data-set and USF Gait Challenge data-set, we achieve recognition rates of 96% and 61% respectively. The lower recognition rate for USF data-set is attributed to both the larger number of subjects and the longer distance from camera to subjects. This suggests that proper zoom lenses are needed to ensure that the gait motion is seen at sufficient detail.

The experimental results show that the sagittal plane contains identification information. In our future work, we would combine the frontal view of the subject for other information such as the toe-out and the bending of the legs [20], to further improve the recognition rate.

Acknowledgments

The authors would like to thank Shan Lu for fruitful discussions, and Stratos Loukidis for setting up the data-base. This work is supported by the National Science Foundation, contract number NSF-ITR-0205671, NSF-ITR-0313184, NSF:0200983, and research scientist funds by the Gallaudet Research Institute..

References

[1] S. Ayer and H. S. Sawhney. Layered representation of motion video using robust maximum-likelihood estimation of mixture models and MDL encoding. In *IEEE International Conference on Computer Vision*, pages 777–784, 1995.

[2] A. Borghese, L. Bianchi, and F. Lacquaniti. Kinematic determinants of human locomotion. *J. Physiology*, 494(3):863–879, 1996.

[3] R. T. Collins, R. Gross, and J. Shi. Silhouette-based human identification from body shape and gait. In *International Conference on Automatic Face and Gesture Recognition*, 2002.

[4] D. Cunado, M.S. Nixon, and J.N. Carter. Using gait as a biometric, via phase-weighted magnitude spectra. In *1st Int. Conf. audio and video based biometric person authentification*, 1997.

[5] A. Elgammal, D. Harwood, and L. Davis. Non-parametric model for background subtraction. In *6th European Conference on Computer Vision*, 2000.

[6] J. P. Foster, M. S. Nixon, and A. Prudel-Bennett. Automatic gait recognition using area-based metrics. *Pattern Recognition Letters*, 24(14):2489–2497, 2001.

[7] R. Gross and J. Shi. The CMU Motion of Body (MoBo) database. Technical Report CMU-RI-TR-01-18, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, June 2001.

[8] J.B. Hayfron-Acquah, M.S. Nixon, and J.N. Carter. Automatic gait recognition by symmetry analysis. *Pattern Recognition Letters*, 24(13):2175–2183, September 2003.

[9] B. K. P. Horn, editor. *Robot Vision*. MIT Press, 1986.

[10] G. Johansson. Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics*, 14(2):201–211, 1973.

[11] A. Kale, N. Cuntoor, B. Yegnanarayana, A.N Rajagopalan, and R. Chellappa. Gait analysis for human identification. In *Proceedings of the 3rd International conference on Audio and Video Based Person Authentication*, 2003.

[12] L. Lee, G. Dalley, and K. Tieu. Learning pedestrian models for silhouette refinement. In *International Conference on Computer Vision and Pattern Recognition*, 2003.

[13] L. Lee and W.E.L. Grimson. Gait analysis for recognition and classification. In *IEEE Conference on Face and Gesture Recognition*, pages 155–161, 2002.

[14] L. Little and J. Boyd. Recognizing people by their gait: the shape of motion. *Videre*, 1(2):1–32, 1996.

[15] S.A. Niyogi and E.H. Adelson. Analyzing and recognizing walking figures in xyt. In *IEEE Conference on Computer Vision and Pattern Recognition*, 1994.

[16] P.J. Phillips, S. Sarkar, I. Robledo, P. Grother, and K. Bowyer. The gait identification challenge problem: Data sets and baseline algorithm. In *International Conference on Pattern Recognition*, 2002.

[17] L. R. Rabiner. A tutorial on Hidden Markov Models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[18] C. Sminchisescu and B. Triggs. Covariance scaled sampling for monocular 3d body tracking. In *IEEE International Conference on Computer Vision and Pattern Recognition*, 2001.

[19] S. V. Stevenage, M. S. Nixon, and K. Vince. Visual analysis of gait as a cue to identity. *Applied Cognitive Psychology*, 13:513–526, 1999.

[20] H. Sun. *Curved Path Human Locomotion on Uneven Terrain*. PhD thesis, University of Pennsylvania, 2000.

[21] A.R. Tilley, editor. *The Measure of Man and Woman: Human Factors in Design*. H.D. Associates, NY, 1993.

Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW'04)
1063-6919/04 $ 20.00 IEEE