# Gait Recognition by Combining Classifiers Based on Environmental Contexts

Ju Han and Bir Bhanu

Center for Research in Intelligent Systems
University of California, Riverside, CA 92521, USA
{bhanu,jhan}@cris.ucr.edu

**Abstract.** Human gait properties can be affected by various environmental contexts such as walking surface and carrying objects. In this paper, we propose a novel approach for individual recognition by combining different gait classifiers with the knowledge of environmental contexts to improve the recognition performance. Different classifiers are designed to handle different environmental contexts, and context specific features are explored for context characterization. In the recognition procedure, we can determine the probability of environmental contexts in any probe sequence according to its context features, and apply the probabilistic classifier combination strategies for the recognition. Experimental results demonstrate the effectiveness of the proposed approach.

## 1 Introduction

Current image-based individual human recognition methods, such as fingerprints, face or iris biometric modalities, generally require a cooperative subject, views from certain aspects and physical contact or close proximity. These methods can not reliably recognize non-cooperating individuals at a distance in the real world under changing environmental conditions. Gait, which concerns recognizing individuals by the way they walk, is a relatively new biometric without these disadvantages. However, gait also has some limitations, it can be affected by clothing, shoes, or other environmental contexts. Moreover, special physical conditions such as injury can also change a person's walking style. The large gait variation of the same person under different conditions (intentionally or unintentionally) reduces the discriminating power of gait as a biometric and it may not be as unique as fingerprint or iris, but the inherent gait characteristic of an individual still makes it irreplaceable and useful in many visual surveillance applications.

In traditional biometric paradigms, individuals of interest are represented by their biometric examples in the gallery data. In general, the gallery examples are obtained under the similar environmental condition (context) and the number of examples for each individual is limited to one. This setup is good for strong biometrics such as iris and fingerprint, where the inherent discriminating features are abundance. Even if the context changes, there are still enough features to distinguish one individual from others.

This setup may not be appropriate for gait recognition. Human gait properties can be affected by various environmental contexts such as walking surface, carrying objects and environmental temperature. The change of an environmental context may introduce a large appearance change in the detected human silhouette, which may lead to

T. Kanade, A. Jain, and N.K. Ratha (Eds.): AVBPA 2005, LNCS 3546, pp. 416–425, 2005.

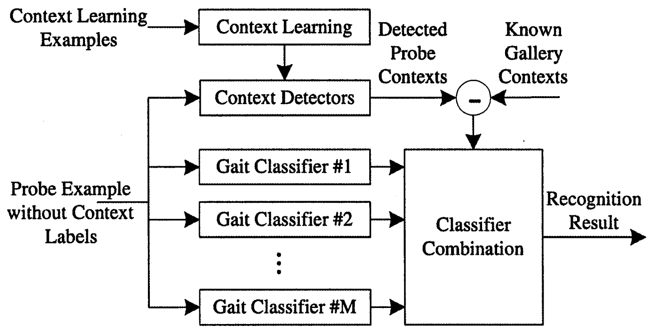


**Fig. 1.** Context-based classifier combination

a failure in recognition. The large gait variation of the same individual under different contexts requires more gallery examples of all individuals from all possible different environmental contexts. However, this requirement is unreal due to the complexity of real-world situations. Due to the difficulty of gait data acquisition, gait gallery examples are generally obtained under one or several environmental conditions and the number of examples for each individual is also very limited. Moreover, The environmental contexts are too rich in the real world to be entirely included in a gallery dataset.

Different gait recognition approaches (classifiers) character gait properties from different aspects. It is difficult to find a single classifier to effectively recognize individuals under all environmental contexts without gallery examples from these contexts. One classifier may be insensitive to the change of one context, while another classifier may be insensitive to the change of another context. If we can detect the environmental contexts of a given probe gait example, it is possible to combine these classifier to improve the recognition performance.

In this paper, we propose a context-based human recognition approach by probabilistically combining different gait classifiers under different environmental contexts. The basic idea is illustrated in Figure 1. First, context properties are learned from context training examples to construct context detectors. The contexts of a given probe gait examples are then obtained by these context detectors. Assuming that all gait gallery examples are obtained under the similar environmental contexts, the context changes between the probe example and gallery examples are obtained. With the gait classifiers designed for individual recognition under different environmental context changes, these classifiers are probabilistically combined to recognize the probe individual based on the detected context changes.

## 2 Related Work

In recent years, various approaches have been proposed for human recognition by gait. These approaches can be divided into two major categories: model-based approaches and model-free approaches.

Model-based gait recognition approaches focus on recovering a structural model of human motion. Niyogi and Adelson [1] find the bounding contours of the walker, and fit a simplified stick model on them. A characteristic gait pattern in spatiotemporal volume is generated from the model parameters for recognition. Yoo et al. [2] estimate hip and knee angles from body contour by linear regression analysis. Then trigonometric-polynomial interpolant functions are fitted to the angle sequences, and the parameters so-obtained are used for recognition. Bhanu and Han [3] propose a kinematic-based approach to recognize individuals by gait. The 3D human walking parameters are estimated by performing a least squares fit of the 3D kinematic model to the 2D silhouette extracted from a monocular image sequence. Human gait signatures are generated by selecting features from the estimated parameters.

Model-free approaches make no attempt to recover a structural model of human motion. Little and Boyd [4] describe the shape of the human motion with a set of features derived from moments of a dense flow distribution. Shutler et al. [5] include velocity into the traditional moments to obtain the so-called velocity moments (VMs). BenAbdelkader et al. [6] use height, stride and cadence for to identify human. Sundaresan et al. [7] proposed a hidden Markov models (HMMs) based framework for individual recognition from their gait. Huang et al. [8] propose a template matching approach by combining transformation based on canonical analysis, with eigenspace transformation for feature selection. Similarly, Wang et al. [9] generate boundary distance vector from the original human silhouette contour as the template, which is used for gait recognition via eigenspace transformation. Phillips et al. [10] propose a direct template matching approach to measure the similarity between the gallery and probe sequences by computing the correlation of corresponding time-normalized frame pairs. Similarly, Collins et al. [11] first extract key frames from a sequence, and the similarity between two sequences is computed from the normalized correlation on key frames only. Tolliver and Collins [12] cluster human silhouettes of each training sequence into $k$ prototypical shapes. Silhouettes in a testing sequence are also classified into $k$ prototypical shapes that are used to compare with those in training sequences.

## 3 Technical Approach

In this section, we describe the proposed context-based classifier combination for individual recognition by gait. The context investigated in this paper is the walking surface type, but the approach could be extended to other contexts. The system diagram is shown in Figure 2.

### 3.1 Gait Representation

We assume that silhouettes have been extracted from original human walking sequences. A silhouette preprocessing procedure [10] is then applied on the extracted silhouette sequences. It includes size normalization (proportionally resizing each silhouette image so that all silhouettes have the same height) and horizontal alignment (centering the upper half silhouette part with respect to its horizontal centroid). In a preprocessed silhouette sequence, the time series signal of lower half silhouette size from each frame indicates the gait frequency and phase information. We estimate the gait frequency and phase by maximum entropy spectrum estimation [4] from the obtained time series signal.

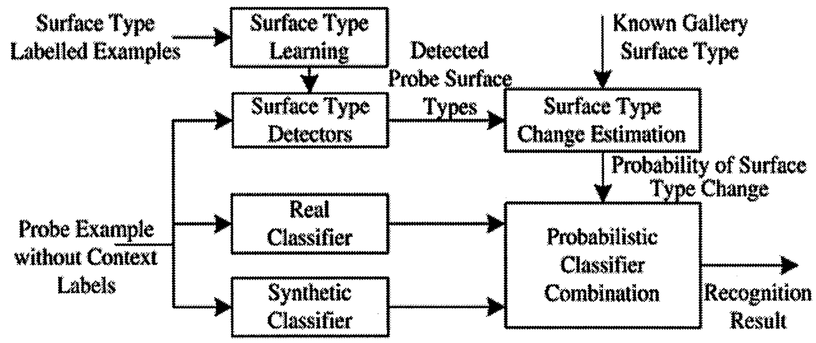


**Fig. 2.** Diagram of context-based classifier combination for individual recognition by gait. The context investigated in this diagram is the walking surface type

Given the preprocessed binary gait silhouette images $B_t(x, y)$ at time $t$ in a sequence, the grey-level gait energy image (GEI) is defined as follows [13]

$$G(x, y) = \frac{1}{N} \sum_{t=1}^{N} B_t(x, y) \tag{1}$$

where $N$ is the number of frames in the complete cycle(s) of a silhouette sequence, $t$ is the frame number in the sequence (moment of time), $x$ and $y$ are values in the 2D image coordinate. It reflects major shapes of silhouettes and their changes over the gait cycle. We refer to it as gait energy image because: (a) each silhouette image is the space-normalized energy image of human walking at this moment; (b) GEI is the time-normalized accumulative energy image of human walking in the complete cycle(s); (c) a pixel with higher intensity value in GEI means that human walking occurs more frequently at this position (i.e., with higher energy). In comparison with binary silhouette sequence, GEI representation saves both storage space and computation time for recognition and is less sensitive to silhouette noise in individual frames. We use GEI as the gait representation for individual recognition in this paper.

### 3.2 Walking Surface Type Detection

Various environmental contexts have effect on silhouette appearance: clothing, shoes, walking surface, camera view, carrying object, time, etc. Among these contexts, slight camera view changes may be neglected. Irregular changes in clothing, shoe, carrying object and time generally cannot be detected. When the same person walks on different surface types, the detected silhouettes may have large difference in appearance. For example, silhouettes on the grass surface may miss the bottom part of feet, while silhouettes on the concrete surface may contain additional shadows. In these cases, silhouette size normalization errors occur, and silhouettes so-obtained may have different scales with respect to silhouettes on other surfaces. Figure 3 shows the GEI examples of three people walking on grass or concrete surfaces in USF HumanID database.

Considering the lower body part difference in silhouettes of people walking on grass and concrete surface, we use the silhouette energy in the lower body part as the indicator

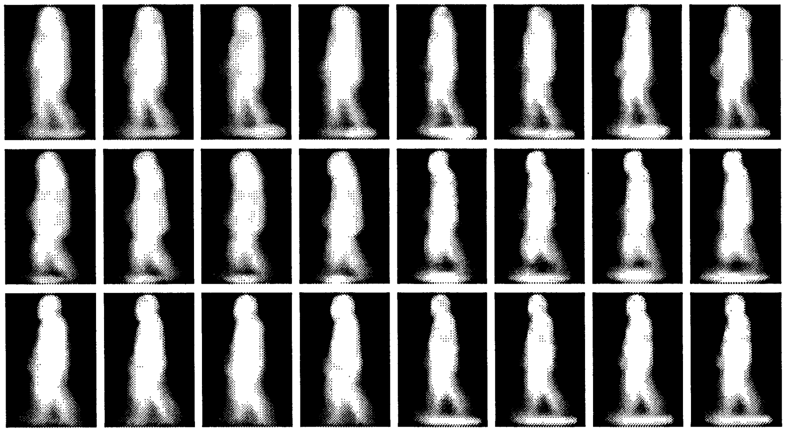

**Fig. 3.** GEI examples of three people (rows) waling on different surface types. First four examples in each row are on the grass surface, and the others are on the concrete surface

of the walking surface type. Let the bottom row be the first row and leftmost column be the first column in the image coordinate, the surface type indicator is defined as

$$s(G, N_{TOP}) = \frac{\sum_{i=1}^{N_{TOP}} \sum_{j=1}^{N_{COL}} G(i,j)}{\sum_{i=1}^{N_{ROW}} \sum_{j=1}^{N_{COL}} G(i,j)}, \tag{2}$$

where $G$ is a GEI example with the size of $N_{ROW} \times N_{COL}$, and $N_{TOP}$ is the number of rows from the bottom. Assuming $s$ has a Gaussian distribution for both grass GEI examples and concrete GEI examples, the class-conditional probability functions are estimated from the context training examples as follows

$$\begin{aligned} p(s|grass) &= \frac{1}{\sqrt{2\pi}\sigma_{grass}} \exp\left\{-\frac{(s-\mu_{grass})^2}{2\sigma_{grass}^2}\right\} \\ p(s|concrete) &= \frac{1}{\sqrt{2\pi}\sigma_{concrete}} \exp\left\{-\frac{(s-\mu_{concrete})^2}{2\sigma_{concrete}^2}\right\} \end{aligned} \tag{3}$$

where $\mu_{grass}$ and $\sigma_{grass}$ are the sample mean and sample standard deviation of $s$ for training examples on the grass surface, and $\mu_{concrete}$ and $\sigma_{concrete}$ are the sample mean and sample standard deviation of $s$ for training examples on the concrete surface. These distributions are different for different $N_{TOP}$ values. The optimal $N_{TOP}$ for discriminating these two surface types is estimated by maximizing the Bhattacharyya distance with respect to $N_{TOP}$:

$$B = \frac{(\mu_{grass} - \mu_{concrete})^2}{4(\sigma_1^2 + \sigma_2^2)} + \frac{1}{2} \ln \frac{\sigma_1^2 + \sigma_2^2}{2\sigma_1 \sigma_2}. \tag{4}$$

The Bhattacharyya distance is used as a class seperability measure here. The Bhattacharyya distance of the two distribution with respect to different $N_{TOP}$ values is shown in Figure 4(a). The estimated distribution for optimal $N_{TOP} = 6$ is shown in Figure 4(b).

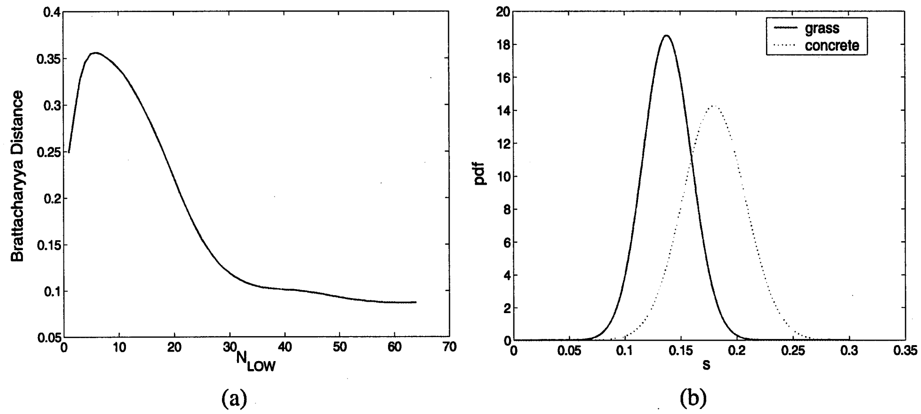


**Fig. 4.** (a) The Bhattacharyya distance of the two distribution with respect to different $N_{TOP}$ values. (b) The estimated distributions of $p(s|grass)$ and $p(s|concrete)$ for $N_{TOP} = 6$

According to the Bayes rule, we have the following probabilities for probabilistic classifier combination

$$P(grass|s) = \frac{p(s|grass)P(grass)}{p(s)}$$
$$P(concrete|s) = \frac{p(s|concrete)P(concrete)}{p(s)}. \tag{5}$$

### 3.3 Classifier Design

In this paper, we use the real gait classifier for recognizing probe examples having no surface type change with respect to gallery examples, and synthetic gait classifier for recognizing probe examples having the surface type change [13].

The real GEI templates for an individual are directly computed from each cycle of the silhouette sequence of this individual. They are used as the input of real classifier for recognizing probe examples having no surface type change with respect to gallery examples.

A statistical feature extraction method by combining PCA and MDA is used for learning real gait features from training real templates. Let $\mathrm{m}_{ri}$ be the mean of real feature vectors belonging to the $i$th class (individual) in the gallery set. Given a probe example $P$, $\{R_j\}$, $j = 1, \ldots, n$, are its real gait templates. The corresponding real feature vector set is obtained as follows

$$\{\hat{R}_P\}: \quad \hat{\mathrm{r}}_j = T_r R_j, \quad j = 1, ..., n$$

where $T_r$ is the learned transformation matrix for real feature extraction. The dissimilarity between the probe example and each gallery class is then measured by

$$D(\hat{R}_P, \omega_i) = \frac{1}{n}\sum_{j=1}^{n} ||\hat{\mathrm{r}}_j - \mathrm{m}_{ri}||, \quad i = 1, ..., c \tag{6}$$

where $c$ is the number of classes in the gallery set. The real classifier is

$$\text{Decide} \quad P \in \omega_k \quad \text{if} \quad D(\hat{R}_P, \omega_k) = \min_{i=1}^{c} D(\hat{R}_P, \omega_i). \tag{7}$$

Although real gait templates provide cues for individual recognition, all the templates from the same sequence are obtained under the "same" physical conditions. If the condition changes, the learned features may not work well for recognition. Let $R_0$ be the GEI template computed all cycles of a given silhouette sequence. Assume that $k$ bottom rows of $R_0$ are missed due to some kind of environmental conditions. According to the silhouette preprocessing procedure in Section 3.1, the remaining part needs to be proportionally resized to fit to the original height. In the same way, we can generate a series of new synthetic GEI templates corresponding to different lower body part distortion with the different values of $k$. The synthetic templates expanded from the same $R_0$ have the same global shape properties but different bottom parts and different scales. Therefore, they provide cues for individual recognition that are less sensitive to surface type changes.

A similar statistical feature extraction method by combining PCA and MDA is used for learning synthetic gait features from synthetic templates. Let $\mathrm{m}_{si}$ be the mean of synthetic feature vectors belonging to the $i$th class (individual) in the gallery set. Given a probe example $P$, $\{S_j\}$, $j = 1, \ldots, m$, are its synthetic gait templates. The corresponding synthetic feature vector set is obtained as follows

$$\{\hat{S}_P\}: \quad \hat{\mathrm{s}}_j = T_r S_j, \quad j = 1, ..., m$$

where $T_s$ is the learned transformation matrix for synthetic feature extraction. The dissimilarity between the probe example and each gallery class is then measured by

$$D(\hat{S}_P, \omega_i) = \frac{1}{m} \sum_{j=1}^{m} ||\hat{\mathrm{s}}_j - \mathrm{m}_{si}||, \quad i = 1, ..., c \tag{8}$$

where $c$ is the number of classes in the gallery set. The synthetic classifier is

$$\text{Decide} \quad P \in \omega_k \quad \text{if} \quad D(\hat{S}_P, \omega_k) = \min_{i=1}^{c} D(\hat{S}_P, \omega_i). \tag{9}$$

### 3.4 Probabilistic Classifier Combination

Given a probe example, the probabilities of different surface types are obtained in Equation (5). The dissimilarities of the probe example of each class in the gallery set are obtained in Equation (6) and (8), respectively. Notice that the real classifier is designed for recognizing probe examples having no surface type change with respect to gallery examples, and the synthetic gait classifier is designed for recognizing probe examples having the surface type change. If walking surface of gallery examples are grass, the combined dissimilarly is measures as follows

$$\begin{aligned} D(P, \omega_i) &= P(grass|s)\bar{D}(\hat{R}_P, \omega_i) + P(concrete|s)\bar{D}(\hat{S}_P, \omega_i) \\ &= P(grass|s)\frac{D(\hat{R}_P, \omega_i)}{\sum_{j=1}^{c} D(\hat{R}_P, \omega_j)} + P(concrete|s)\frac{D(\hat{S}_P, \omega_i)}{\sum_{j=1}^{c} D(hatS_P, \omega_j)} \end{aligned}$$

**Table 1.** Twelve experiments designed for human recognition in USF HumanID database

| Experiment Label | Size of Probe Set | Difference between Gallery and Probe Sets |
|---|---|---|
| A | 122 | View |
| B | 54 | Shoe |
| C | 54 | View and Shoe |
| D | 121 | Surface |
| E | 60 | Surface and Shoe |
| F | 121 | Surface and View |
| G | 60 | Surface, Shoe and View |
| H | 120 | Briefcase |
| I | 60 | Shoe and Briefcase |
| J | 120 | View and Briefcase |
| K | 33 | Time, Shoe and Clothing |
| L | 33 | Surface and Time |

for $i = 1, ..., c$, where $\bar{D}$ is the normalized dissimilarity. Assuming $P(grass) = P(concrete)$, we have

$$D(P, \omega_i) = P(s|grass)\frac{D(\hat{R}_P, \omega_i)}{\sum_{j=1}^{c} D(\hat{R}_P, \omega_j)} + P(s|concrete)\frac{D(\hat{S}_P, \omega_i)}{\sum_{j=1}^{c} D(\hat{S}_P, \omega_j)} \quad (10)$$

for $i = 1, ..., c$. The combined classifier based on surface context is

$$\text{Decide} \quad P \in \omega_k \quad \text{if} \quad D(P, \omega_k) = \min_{i=1}^{c} D(P, \omega_i). \quad (11)$$

## 4 Experimental Results

Our experiments are carried out on the USF HumanID gait database [10]. This database consists of people walking in elliptical paths in front of the camera. For each person, there are up to 5 covariates: viewpoints (left/right), shoe types (A/B), surface types (grass/concrete), carrying conditions (with/without a briefcase), and time and clothing. Twelve experiments are designed for individual recognition as shown in Table 1. The gallery set contains 122 sequences. Individuals are unique in the gallery and each probe set, and there are no common sequence among the gallery set and all probe sets. The walking surface type in the gallery set is grass.

Phillips et al. [10] propose a baseline approach to extract human silhouette and recognize people in this database. For comparison, they provide extracted silhouette data which can be found at the website http://marathon.csee.usf.edu/GaitBaseline/. Our experiments begin with these extracted binary silhouette data (version 2.1) that are updated on September 5, 2003. The performance of their baseline algorithm are shown in Table 2. In this table, rank1 means that only the first subject in the retrieval rank list is recognized as the same subject as the query subject, and rank5 means that the first five subjects are all recognized as the same subject as the query subject. The performance in the table is the recognition rate under these two definitions.

**Table 2.** Comparison of recognition performance among different approaches on silhouette sequence version 2.1 (Legends: baseline - USF baseline algorithm [10]; real - real classifier; synthetic - synthetic classifier; context - proposed context-based approach, this paper)

| | Rank1 Performance | | | | Rank5 Performance | | | |
|---|---|---|---|---|---|---|---|---|
| | baseline | real | synthetic | context | baseline | real | synthetic | context |
| A | 73% | 89% | 84% | 90% | 88% | 93% | 93% | 93% |
| B | 78% | 87% | 93% | 91% | 93% | 93% | 96% | 94% |
| C | 48% | 78% | 67% | 80% | 78% | 89% | 93% | 89% |
| D | 32% | 36% | 53% | 56% | 66% | 65% | 75% | 81% |
| E | 22% | 38% | 55% | 57% | 55% | 60% | 71% | 76% |
| F | 17% | 20% | 30% | 27% | 42% | 42% | 53% | 53% |
| G | 17% | 28% | 34% | 36% | 38% | 45% | 53% | 50% |
| H | 61% | 62% | 47% | 60% | 85% | 87% | 79% | 90% |
| I | 57% | 59% | 57% | 62% | 78% | 79% | 81% | 84% |
| J | 36% | 58% | 40% | 57% | 62% | 81% | 65% | 84% |
| K | 3% | 3% | 21% | 9% | 3% | 6% | 33% | 18% |
| L | 3% | 6% | 24% | 12% | 15% | 9% | 42% | 27% |

We carry out experiments of human recognition by the real classifier, the synthetic classifier and the combined classifier based context according to rules in (7), (9), and (10), respectively. Table 2 shows the recognition performance of USF baseline algorithm and our proposed approaches. Note that the rank1 and rank5 performance of proposed classifiers is better than or equivalent to that of baseline algorithm on all experiments.

The performance of the synthetic classifier is significantly better than that of the real classifier on experiments D-G and L, where the surface type of probe examples is different from that of gallery examples. In other experiments where the surface type of probe examples is the same as that of gallery examples, the performance of the real classifiers is better on A, C and G-J, but a little worse on B and K. These results demonstrate the our designed real and synthetic classifiers is suitable for their desired contexts.

The combined classifier based on the surface context achieves better performance than individual real feature classifier and synthetic classifier in most experiments. It is shown that the combined classifier takes advantage of merits in individual classifiers based on the detected context information. In this paper, we only detect and use the specific context information about the walking surface type, and only design two classifiers for it. If we can detect or obtain more context information such as carrying objects, clothing and time, and design the corresponding classifiers, we expect further improved combination results.

## 5 Conclusions

In this paper, we propose a context-based human recognition approach by probabilistically combining different gait classifiers with different environmental contexts. First, context properties are learned from context training examples to construct context detectors. The contexts of a given probe gait examples are then obtained by these context detectors. With the gait classifiers designed for individual recognition under different environmental context changes, these classifiers are probabilistically combined to recognize the probe individual based on the detected context changes. Experimental results show that the combined classifier takes advantage of merits in individual classifiers based on the detected context information.

## References


1. Niyogi, S., Adelson, E.: Analyzing and recognizing walking figures in xyt. Proc. IEEE Conference on CVPR (1994) 469–474
2. Yoo, J.H., Nixon, M., Harris, C.: Model-driven statistical analysis of human gait motion. Proc. IEEE International Conference on Image Processing **1** (2002) 285–288
3. Bhanu, B., Han, J.: Individual recognition by kinematic-based gait analysis. Proc. International Conference on Pattern Recognition **3** (2002) 343–346
4. Little, J., Boyd, J.: Recognizing people by their gait: the shape of motion. Videre: Journal of Computer Vision Research **1** (1998) 1–32
5. Shutler, J., Nixon, M., Harris, C.: Statistical gait recognition via velocity moments. Proc. IEE Colloquium on Visual Biometrics (2000) 10/1–10/5
6. BenAbdelkader, C., Cutler, R., Davis, L.: Person identification using automatic height and stride estimation. Proc. International Conference on Pattern Recognition **4** (2002) 377–380
7. Sundaresan, A., RoyChowdhury, A., Chellappa, R.: A hidden Markov model based framework for recognition of humans from gait sequences. Proc. ICIP **2** (2003) 93–96
8. Huang, P., Harris, C., Nixon, M.: Recognizing humans by gait via parameteric canonical space. Artificial Intelligence in Engineering **13** (1999) 359–366
9. Wang, L., Hu, W., Tan, T.: A new attempt to gait-based human identification. Proc. International Conference on Pattern Recognition **1** (2002) 115–118
10. Phillips, P., Sarkar, S., Robledo, I., Grother, P., Bowyer, K.: The gait identification challenge problem: data sets and baseline algorithm. Proc. ICPR **1** (2002) 385–388
11. Collins, R., Gross, R., Shi, J.: Silhouette-based human identification from body shape and gait. Proc. IEEE Intl. Conf. on Automatic Face and Gesture Recognition (2002) 351–356
12. Tolliver, D., Collins, R.: Gait shape estimation for identification. Proc. 4th International Conference on Audio- and Video-Based Biometric Person Authentication, LNCS 2688 (2003) 734–742
13. Han, J., Bhanu, B.: Statistical feature fusion for gait-based human recognition. Proc. IEEE Conf. on Computer Vision and Pattern Recognition **2** (2004) 842–847