Gait Recognition Using Gait Entropy Image

Khalid Bashir, Tao Xiang, Shaogang Gong

School of Electronic Engineering and Computer Science,
Queen Mary University of London, United Kingdom
{khalid,txiang,sgg}@dcs.qmul.ac.uk

Abstract

Gait as a behavioural biometric is concerned with how people walk. However, most existing gait representations capture both motion and appearance information. They are thus sensitive to changes in various covariate conditions such as carrying and clothing. In this paper, a novel gait representation termed as Gait Entropy Image (GEnI) is proposed. Based on computing entropy, a GEnI encodes in a single image the randomness of pixel values in the silhouette images over a complete gait cycle. It thus captures mostly motion information and is robust to covariate condition changes that affect appearance. Extensive experiments on the USF HumanID dataset, CASIA dataset and the SOTON dataset have been carried out to demonstrate that the proposed gait representation outperforms existing methods, especially when there are significant appearance changes. Our experiments also show clear advantage of GEnI over the alternatives without the assumption on cooperative subjects, i.e. both the gallery and the probe sets consist of a mixture of gait sequences under different and unknown covariate conditions.

1 Introduction

Gait is a biometric which measures the way people walk. Among various image-based biometrics, face, iris, fingerprint, and gait are the most widely studied ones. Gait is in general perceived as a weaker biometric compared to the others. This is largely because gait as a behavioural biometrics is more likely to be affected by changes in covariate conditions. Some of these conditions mainly affect the appearance of people such as clothing condition, carrying condition and view angle, while others affect gait itself such as shoes, time and surface. Despite of this limitation, gait has it distinctive strength in applications such as visual surveillance because it does not require cooperative subjects viewed at a close proximity whilst other image-based biometrics do.

Based on the way gait is represented the existing gait recognition approaches can be divided into two categories: model based and model free approaches. Model based approaches [17, 10, 16, 2] represent gait using the parameters of a model of the body configuration which are estimated over time. This process is usually computationally intensive and error prone

and also requires high resolution images as input. Due to these limitations the current trend on gait representation seems to favour model free approaches. Model free approaches construct gait descriptors from the silhouettes extracted from the video sequences. One of the most adopted model free representation is Gait Energy Image (GEI) [3], which represents gait using a single grey scale image obtained by averaging the silhouettes extracted over a complete gait cycle. A representation that is similar to GEI is Motion Silhouette Image (MSI) [6], where the intensity value of a pixel is computed as a function of motion of that pixel in the temporal dimension over all silhouettes from a gait cycle. Both the GEI and MSI are compact representation of gait (a gait cycle is represented using a single image), which are easy to compute and insensitive to noise in silhouette extraction. However, they are vulnerable to appearance changes of the human silhouette. Although recent studies suggest that static shape information is more important than kinematics for most of the silhouette-based gait recognition approaches [13, 14], including static appearance features in gait representation is a double edged sword and may not be the best choice under covariate condition changes. To overcome this problem, Shape Variation Based (SVB) Frieze Pattern is proposed in [9, 8]. Frieze pattern represents the information contained in a gait sequence by horizontal and vertical projections of the silhouettes. Its extension, SVB Frieze patterns use key frame subtraction in order to mitigate the affects of appearance changes on the silhouette. Although it has been shown that SVB Frieze pattern outperforms GEI and MSI when there are significant appearance changes, its performance is inferior when no changes are present. In addtion, compared to GEI and MSI, it requires a non-trivial temporal alignment process for each gait cycle and is computationally more expensive.

In this paper we propose a new gait representation which is termed as Gait Entropy Image (GEnI). Based on computing entropy, a GEnI encodes in a single image the randomness of pixel values in the silhouette images over a complete gait cycle. Dynamic body areas which undergo consistent relative motion during a gait cycle (e.g. leg, arms) will lead to high gait entropy value, whereas those areas that remain static (e.g. torso) would give rise to low values. A GEnI thus captures mostly motion information and is robust to covariate condition changes that affect the static areas of human body. Compared with existing model free gait representations, GEnI shares the strength of GEI and MSI (i.e. compact, easy to compute and robust to image noise), and the strength of SVB Frieze Pattern (i.e. ro-

bust to covariate condition changes). In the meantime, it avoids the aforementioned weakness of existing representations.

We demonstrate the effectiveness of our GEnI representation through extensive experiments on three public benchmarking datesets: the CASIA gait database [15], the HumanID USF database [11] and the Southampton SOTON database [12]. Our results suggest that the proposed gait representation outperforms existing methods, especially when there are significant appearance changes. Our experiments also show clear advantage of GEnI over the alternatives without the assumption on cooperative subjects [1], i.e. both the gallery and the probe sets consist of a mixture of gait sequences under different and unknown covariate conditions.

2 Gait Entropy Image

Gait Entropy Image (GEnI) is computed from normalised silhouettes. First, silhouettes are extracted using background subtraction from each image frame [11]. Second, the height of the silhouettes are normalised which is followed by the centre alignment. Gait cycles are then estimated using the maximum entropy estimation method in the lower half of the image [11]. Fig. 1 shows examples of normalized silhouettes from the USF, CASIA and SOTON Datasets. The silhouettes from the indoor CASIA and SOTON datasets are clean, whereas the silhouette for the outdoor USF dataset is much more noisy as expected.
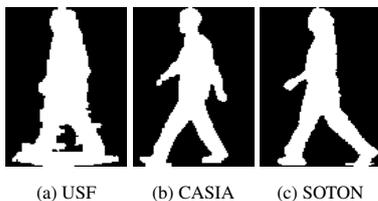

(a) USF   (b) CASIA   (c) SOTON

Figure 1. Example normalised silhouettes from USF, CASIA and SOTON Datasets.

Given a gait cycle of size-normalised and centre-aligned silhouettes, a GEnI is computed by calculating Shannon entropy for each pixel in the silhouette images. Shannon entropy measures the uncertainty associated with a random variable. Considering the intensity value of the silhouettes at a fixed pixel location as a discrete random variable, the entropy of this variable over a complete gait cycle can be computed as

$$H(x,y) = -\sum_{k=1}^{K} p_k(x,y) \log_2 p_k(x,y), \qquad (1)$$

where x, y are the pixel coordinates and $p_k(x,y)$ is the probability that the pixel takes on the kth value. In our case the silhouettes are binary images and we thus have $K = 2$. A Gait Entropy Image $G(x,y)$ can then be obtained by scaling and discretising $H(x,y)$ so that its value ranges from 0 to 255 as follows.

$$G(x,y) = \frac{(H(x,y) - H_{min}) * 255}{(H_{max} - H_{min})}, \qquad (2)$$

where $H_{min} = min(H(x,y))$ and $H_{max} = max(H(x,y))$. Note that since the Gait Entropy Image is computed using the frames in a complete gait cycle there is no temporal alignment problem.

(a) Normal   (b) Bag   (c) Coat

(d) Left View   (e) Right View   (f) Briefcase
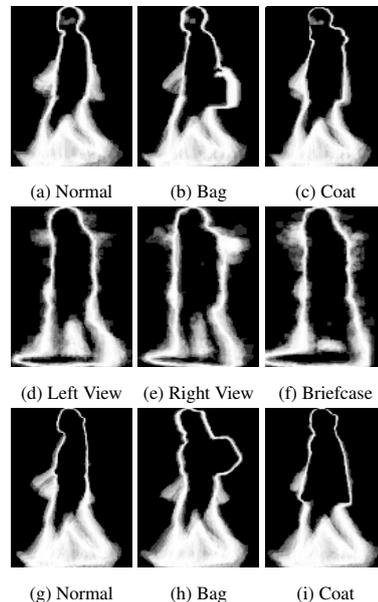
(g) Normal   (h) Bag   (i) Coat

Figure 2. Examples of Gait Entropy Images from the three datasets. Top row CASIA, middle USF and SOTON bottom row

Fig. 2 shows some examples of Gait Entropy Images from the three gait datasets. It clearly shows that the dynamic area of human body, including legs and arms which undergo motions in relation to other body parts, are represented by higher intensity values in the GEnIs. This is because silhouette pixel values in the dynamic areas are more uncertain and thus more informative leading to higher entropy values. In contrast, the static areas such as torso give rise to low intensity values. Importantly, it can be seen from Fig. 2 that the effect of appearance related changes caused by the changes in carrying and clothing condition is significantly reduced and is only shown in the outer contour of the human body. It can also be seen from the GEnIs from the outdoor USF dataset (Fig. 2(d)-(f)) that GEnI is effective in minimising the effects of shadows. In addition, Fig. 2 shows that the outer contour of the human body which contains useful information for distinguishing different people is preserved.

The characteristics of GEnI, in comparison with the existing model free gait representations, are highlighted in Fig. 3. In the top row of Fig. 3, four differnt representations of the same gait cycle, namely the proposed Gait Entropy Image (GEnI), Motion Silhouette Image (MSI), Gait Energy Image (GEI), and Shape Variation Based (SVB) Frieze Pattern, are shown. GEnIs and the GEIs of the same person depicted in the top row but under different carrying and clothing conditions are shown in the bottom row. It can be seen from Fig. 3 that the difference (error) between the GEIs of the same person under normal and carrying conditions (see Fig. 3(c) and (f)) is much bigger

(a) GEnI (b) MSI (c) GEI (d) SVB

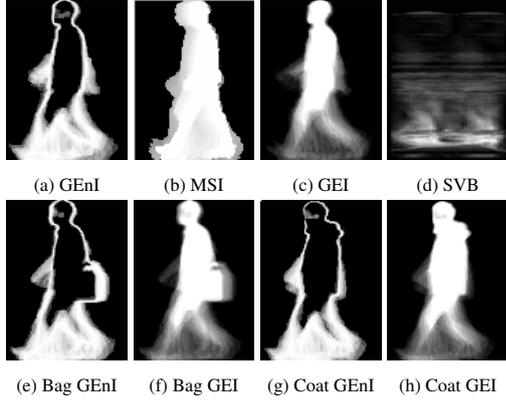(e) Bag GEnI (f) Bag GEI (g) Coat GEnI (h) Coat GEI

Figure 3. GEnI in comparison with alternative model free representations.

than that of GEnIs (see Fig. 3(a) and (e)). This suggests that it would be much harder to recognise that the two GEIs are from the same person. A similar observation can be obtained in Fig. 3 when the clothing condition changes (people wearing bulky coat). MSI would suffer from the same problem of being sensitive to covariate changes as GEI. Those appearance related covariate conditions have less effect on SVB Frieze Pattern as the dynamic areas of human body are given high values (see Fig. 3(d)). However, compared to GEnI, SVB is much harder to compute and needs to address the non-trivial gait cycle temporal alignment problem through key frame selection for each gait cycle. In addition, as indicated by results in [8] and our experiments (see Sec. 4), the performance of SVB is inferior to other representations when there are no covariate changes.

3 Gait Recognition using GEnI

Once gait sequences are represented as GEnIs, gait recognition can be performed by matching a probe GEnI to the gallery GEnI that has the minimal distance to the probe GEnI. We adopt the subspace Component and Discriminant Analysis (CDA). The CDA is based on Principal Component Analysis (PCA) and Multiple Discriminant Analysis (MDA) which seeks to project the original features to a subspace of lower dimensionality so that the best data representation and class separability can be achieved simultaneously [4].

Suppose we have N d-dimensional gallery GEnI templates $\{\mathbf{x}_1, ..., \mathbf{x}_n, ..., \mathbf{x}_N\}$ belonging to c different classes (individuals), where each template is a column vector obtained by concatenating the rows of the corresponding GEnI which are subjected to PCA. PCA is an orthogonal linear transformation that transforms the data to a subspace of dimensionality $\tilde{d}$ (with $\tilde{d} < d$). The PCA subspace keeps the greatest variances by any projection of the data so that the reconstruction error defined below is minimised:

$$J_{\tilde{d}} = \sum_{n=1}^{N} \left\| \left(\mathbf{m} + \sum_{j=1}^{\tilde{d}} a_{nj}\mathbf{e}_j \right) - \mathbf{x}_n \right\|^2 \quad (3)$$

where $\mathbf{m}$ is the mean of the data, $\{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_{\tilde{d}}\}$ are a set of

orthogonal unit vectors representing the new coordinate system of the subspace, a_{nj} is the projection of the nth data to $\mathbf{e}_j$. $J_{\tilde{d}}$ is minimised when $\{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_{\tilde{d}}\}$ are the $\tilde{d}$ eigenvectors of the data covariance matrix with the largest eigenvalues (in decreasing order). Now the gallery template $\mathbf{x}_n$ is represented as a $\tilde{d}$-dimensional feature vector $\mathbf{y}_n$ and we have

$$\mathbf{y}_n = M_{pca}\mathbf{x}_n = [\mathbf{e}_1, ..., \mathbf{e}_{\tilde{d}}]^T \mathbf{x}_n. \quad (4)$$

PCA is followed by MDA which aims to find a subspace where data from different classes are best separated in a least square sense. Different from PCA, MDA is a supervised learning method which requires the gallery data to be labelled into classes. The MDA transformation matrix, W maximises

$$J(W) = \frac{|W^T S_B W|}{|W^T S_W W|}$$

where S_B is the between-class scatter matrix and S_W the within-class scatter matrix of the training (gallery) data in the PCA subspace $\{\mathbf{y}_1, ..., \mathbf{y}_n, ..., \mathbf{y}_N\}$. $J(W)$ is maximised by setting the columns of W to the generalised eigenvectors that correspond to the $c - 1$ nonzero eigenvalues in

$$S_B \mathbf{w}_j = \lambda^j S_W \mathbf{w}_j$$

where $\mathbf{w}_j$ is the jth column of W and c is the number of classes in the training data. Denoting these generalised eigenvectors as $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_{c-1}\}$, a gallery template is represented in the MDA subspace as:

$$\mathbf{z}_n = M_{mda}\mathbf{y}_n = [\mathbf{v}_1, ..., \mathbf{v}_{c-1}]^T \mathbf{y}_n. \quad (5)$$

Note that the choice of $\tilde{d}$ is affected by the dimensionality of the MDA subspace, i.e. $c - 1$. In particular, S_W becomes singular when $\tilde{d} < c$ or $\tilde{d} \gg c$. We therefore set $\tilde{d} = 2c$ in this paper. Now after three steps of dimensionality reduction, both the gallery and probe GEnI feature vectors are represented in a $c - 1$ dimensional subspace and recognition can be performed by computing the distance of the probe feature vector to the gallery feature vectors and match the two with minimal distance.

4 Experiments

We performed comprehensive experiments on the HumanID USF dataset, the CASIA dataset and the Southampton Human ID at a distance (SOTON) datasets to thoroughly evaluate the performance of Gait Entropy Image as gait representation in both outdoor and indoor environments. These three datasets are the largest datasets available for benchmarking gait recognition techniques.

4.1 CASIA Dataset

The CASIA Gait Database [15] is an indoor gait dataset and comprises of 124 subjects. For each subject there are 10 walking sequences consisting of 6 normal walking sequences where the subject does not carry a bag or wear a bulky coat

(CASIASetA), 2 carrying-bag sequences (CASIASetB) and 2 wearing-coat sequences (CASIASetC). Each sequence contains multiple gait cycles. The original image size of the database is 320x240. The first 4 of the 6 normal sequences were used as the gallery set. The probe set included the rest of the normal sequences (CASIASetA2), CASIASetB and CASIASetC.

Probe Set	TM	GEI	GEnI
CASIASetA2	97.6%	**99.4%**	98.3%
CASIASetB	52.0%	60.2%	**80.1%**
CASIASetC	32.7%	30.0%	**33.5%**

Table 1. Comparing different approaches on the CASIA dataset.

The performance of our GEnI representation was measured using recognition rate and compared with GEI with the identical recognition algorithm (see Sec. 3), and a direct template matching based method in [15]. The result is given in Table 1. It can be seen from Table 1 that when the covariate conditions are the same for the probe and gallery set (CASIASetA2), that is without carrying bag and wearing bulking coat, all three methods yield very high recognition rate. However, when the covariate conditions are different, the performance of all three methods are affected. Nevertheless, in this case our GEnI outperforms the other two methods, especially for CASIASetB which corresponds to carrying bag sequences. This indicates that our GEnI gait representation is more robust given changes in covarite conditions.

4.2 SOTON Dataset

The SOTON database [12] consists of two datasets: a large dataset (with 116 subjects) and a small dataset (with 11 subjects). The large dataset consist of 6 subset, numbered as Set A to F. In this experiment, only Set A was used because it is the most widely used and our result can be compared with those reported in the literature. In Set A, all subjects were captured under the normal and fixed covariate conditions. Only one sequence was provided for each subject which contains multiple gait cycles. We used half of the sequences for each subject for training (gallery) and the other half for testing (probe). The small dataset, on the other hand, was designed to investigate the robustness of gait recognition techniques under changing covariate conditions, including carrying and clothing. In the small dataset for each subject one normal sequence (SotonSmallSetA), four carrying-bag sequences (SotonSmallSetB) and one wearing-coat sequences (SotonSmallSetC) were used for testing. The gallery set for the small dataset was the other normal sequences in the small dataset, referred to as SotonSmallSetAN. The original image size for both datasets is 720x576.

The performance of our GEnI was compared with four existing model-free gait representation approaches, namely MSI [6], Frieze Patterns [9], SVB Frieze Patterns [8], and GEI [3]. Table 2 suggests that when the covariate condition is normal

MSI	Frieze	SVB Frieze	GEI	GEnI
84.8%	96.0%	84.0%	99.1	**99.1%**

Table 2. Comparing different approaches on the SOTON large dataset.

Probe Set	GEI	GEnI
SotonSmallSetA	100%	**100%**
SotonSmallSetB	86.3%	**86.3%**
SotonSmallSetC	72.73%	**81.82%**

Table 3. Comparing GEI and GEnI on the SOTON small dataset.

for both the gallery and probe sets, GEI and GEnI outperform the other three approaches. The performance of SVB Frieze patterns was particularly weak compared with our approach, which sacrifices discriminative power for the robustness under covariate condition changes. Since few results are reported in the literature on the SOTON small dataset, we only compare out GEnI with GEI. The result is shown in Table 3. Again, when there is no covariate condition change, both GEnI and GEI achieve very good recognition rate. However, the strength of GEnI becomes clear when the probe set has different covariate conditions which cause appearance changes.

4.3 HumanID USF Dataset

The HumanID USF dataset [11] consists of 122 subjects walking in an elliptical path captured under outdoor conditions. This is one of the most challenging dataset around with variable covariate conditions including carrying briefcase, surface, shoe, view and time.

Probe Set	Baseline	MSCT+SST	GEI	HMM	GEnI
A	73%	80%	89%	89%	89%
B	78%	89%	87%	88%	89%
C	48%	72%	78%	68%	80%
D	32%	14%	36%	35%	30%
E	22%	10%	38%	28%	38%
F	17%	10%	20%	15%	20%
G	17%	13%	28%	21%	22%
H	61%	49%	62%	85%	82%
I	57%	43%	59%	80%	63%
J	36%	30%	59%	58%	66%
K	3%	39%	3%	17%	6%
L	3%	9%	6%	15%	9%
Total	40.9%	38.3%	50.1%	53.5%	**53.5%**

Table 4. Comparing the baseline [11], MSCT+SST [7], GEI [3] and HMM [5] and GEnI on the USF dataset.

For benchmarking purposes 12 experiments A-L have been designed to test the performance of state of the art algorithms. These experiments can be grouped into sets based upon a com-

mon factor such as A-C where the common factor is view; for D-G it is surface; briefcase for H-J and time for K-L. The dataset provides the silhouettes extracted from the video sequences, which were used to compute GEnIs.

Table 4 shows the comparative result obtained on the USF dataset. It can be seen from Table 4 that our approach outperforms the three model-free approaches, namely baseline [11], MSCT+SST [7], and GEI [3], and comparable to a HMM based approach [5]. More specifically, compared with the alternative model-free approaches, GEnI yields much improved result on H-J where subjects carried briefcases. This is consistent with the results we obtained on the indoor CASIA and SOTON datasets. It is also noted that the performance of GEnI is slightly worse than GEI on D and G where the surface was different from that in the gallery set. This is because compared to GEI, the information contained in the dynamic areas is given more weight (see Fig. 3). This characteristic makes GEnI more robust to appearance related covariate changes which affect mainly the static areas. However, when the covariate condition changes such as surface and shoe affect the way people walk (i.e. the gait itself) thus the dynamic areas of human body, our GEnI will also be more sensitive to these changes because of its emphasis on the dynamic areas.

4.4 Evaluation without assuming subject cooperation

In the experiments described so far, the gallery set consist of gait sequences of people under similar covariate conditions. This is essentially to assume that the data was collected in a cooperative manner so that those conditions are known a priori. Nevertheless, this assumption is often invalid in a real-world application where no subject cooperation can be assured. To test the effectiveness of GEnI in an uncooperative environment we put to test our representation under the experimental setup proposed in [1] where the gallery set is composed of a mixture of gait sequences under different unknown covariate conditions. The CASIA and the SOTON small datasets were used in this experiment. For the CASIA dataset we selected the first one third of the sequences from CASIASetC, the second one third from CASIASetB and the last one third from CASIASetA for the gallery set. The probe sets consist of the rest of the dataset and are referred to as CASIASetA2, CASIASetB2 and CASIASetC2. For the SOTON small dataset the mixed gallery set contains the first one third of the subjects from SotonSmallSetAN the second one third from one of the sequences in SotonSmallSetB and last one third from one of the sequences in SotoSmallSetC. The probe set consists of SotonSmallSetA, SotonSmallSetB2 (bag sequences excluding those in the gallery set) and SotonSmallSetC2 (coat sequences excluding those in the gallery set) respectively. The probe and the gallery sets are mutually exclusive.

We compared the performance of GEI and GEnI and the result is shown in Table 5. Comparing Table 5 with Table 1 and Table 3, it is evident that under this more challenging and realistic experimental setup, the performance of both approaches drop significantly. However, our GEnI's performance drops more gracefully and is still superior to that of GEI.

Probe Set	GEI	GEnI
CASIASetA2	48.1%	**48.1%**
CASIASetB2	31.9%	**44.4%**
CASIASetC2	9.7%	**16.1%**
SotonSmallSetA	45.45%	**45.45%**
SotonSmallSetB2	31.82%	**63.64%**
SotonSmallSetC2	36.36%	**36.36%**

Table 5. Comparing GEI with GEnI without assuming subject cooperation.

5 Conclusion

In this paper we have proposed a new gait representation termed Gait Entropy Image (GEnI). GEnI captures mostly motion information and is thus robust to covariate condition changes that affect appearance. Extensive experiments on the USF HumanID dataset, CASIA dataset and the SOTON dataset have been carried out to demonstrate that the proposed gait representation outperforms existing methods, especially when there are significant appearance changes. It has been noted that the performance of GEnI is affected when the covariate conditions affect gait itself. The ongoing work includes the investigation of extracting features from GEnI rather than using it directly as template in order to make it more robust to circumstances where gait itself changes e.g. different shoe, surface and/or time.

References

[1] K. Bashir, T. Xiang, and S. Gong. Feature selection for gait recognition without subject cooperation. In *BMVC*, 2008.

[2] T. Ding. A robust identification approach to gait recognition. In *CVPR*, pages 1–8, 2008.

[3] J. Han and B. Bhanu. Individual recognition using gait energy image. *PAMI*, 28(2):316–322, Feb 2006.

[4] P.S. Huang, C.J. Harris, and M.S. Nixon. Recognizing humans by gait via parametric canonical space. *Artificial Intelligence in Eng.*, 13:359–366, 1999.

[5] A. Kale, A. Sundaresan, A.N. Rajagopalan, N.P. Cuntoor, A.K. Roy Chowdhury, V. Kruger, and R. Chellappa. Identification of humans using gait. *IEEE Transactions on Image Processing*, 13(9):1163–1173, September 2004.

[6] T.H.W. Lam and R.S.T. Lee. A new representation for human gait recognition: Motion silhouettes image (msi). In *ICB*, pages 612–618, 2006.

[7] T.H.W. Lam, R.S.T. Lee, and D. Zhang. Human gait recognition by the fusion of motion and static spatio-temporal templates. *PR*, 40(9):2563–2573, September 2007.

[8] S.K. Lee, Y.X. Liu, and R.T. Collins. Shape variation-based frieze pattern for robust gait recognition. In *CVPR*, pages 1–8, 2007.

[9] Y. Liu, R. Collins, and Y. Tsin. Gait sequence analysis using frieze patterns. In *ECCV*, May 2002.

[10] H. Lu and P. Venetsanopoulos. A layered deformable model for gait analysis. In *AFGR*, pages 249–254, April 2006.

[11] S. Sarkar, P. Phillips, Z. Liu, I. Vega, P. Grother, and K. Bowyer. The humanID gait challenge problem: Data sets, performance, and analysis. *PAMI*, 27(2):162–177, 2005.

[12] J. Shutler, M. Grant, M. Nixon, and J Carter. On a large sequence-based human gait database. In *RASC*, pages 66–71, 2002.

[13] A. Veeraraghavan, A. Chowdhury, and R. Chellappa. Role of shape and kinematics in human movement analysis. In *CVPR*, pages 730–737, June 2004.

[14] G.V. Veres, L. Gordon, J.N. Carter, and M.S. Nixon. What image information is important in silhouette-based gait recognition? In *CVPR*, pages 776–782, June 2004.

[15] S. Yu, D. Tan, and T. Tan. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In *ICPR*, pages 441–444, 2006.

[16] R. Zhang, C. Vogler, and D. Metaxas. Human gait recognition at sagittal plane. *Image Vision Comput.*, 25(3):321–330, 2007.

[17] G. Zhao, G. Liu, H. Li, and Pietikainen. 3D gait recognition using multiple cameras. In *AFGR*, pages 529–534, April 2006.