Gait Energy Volumes and Frontal Gait Recognition using Depth Images

Sabesan Sivapalan, Daniel Chen, Simon Denman, Sridha Sridharan and Clinton Fookes
Image and Video Research Laboratory
Queensland University of Technology
GPO Box 2434, 2 George St.
Brisbane, Queensland 4001
{*sivapalen.sabesan*,*daniel.chen*, *s.denman*,*s.sridharan*,*c.fookes*}@qut.edu.au

Abstract

Gait energy images (GEIs) and its variants form the basis of many recent appearance-based gait recognition systems. The GEI combines good recognition performance with a simple implementation, though it suffers problems inherent to appearance-based approaches, such as being highly view dependent. In this paper, we extend the concept of the GEI to 3D, to create what we call the gait energy volume, or GEV. A basic GEV implementation is tested on the CMU MoBo database, showing improvements over both the GEI baseline and a fused multi-view GEI approach. We also demonstrate the efficacy of this approach on partial volume reconstructions created from frontal depth images, which can be more practically acquired, for example, in biometric portals implemented with stereo cameras, or other depth acquisition systems. Experiments on frontal depth images are evaluated on an in-house developed database captured using the Microsoft Kinect, and demonstrate the validity of the proposed approach.

1. Introduction

The recognition of people is a challenging task in the computer vision field. A number of biometrics such as gait, fingerprint, iris and face are used for this purpose. Gait has a unique advantage over other biometrics in that it can be recognised from a distance without alerting the subject. There are many successful gait recognition techniques that have been developed, however, they still struggle to perform well under certain factors, such as changes in viewing angles.

There are two major approaches to gait recognition; appearance-based (model-free) and model-based [12]. Model-based techniques gather gait dynamics directly by modelling the underlying kinematics of human motion. The trade off though is that these algorithms are generally more complex and computationally expensive. Examples of this approach include the work of Cunado *et al*. [2], where a motion-based model is fitted to an extracted binary silhouette to analyse the angular motion of the hip and thigh by means of a Fourier series. Wagg and Nixon [15] proposed bulk motion and shape estimation guided by bio-mechanical analysis and used mean gait data to create the motion models.

In contrast to these model-based approaches, appearance-based techniques are less complex to implement and establish correspondence between successive frames based upon the implicit notion of what is being observed. Earlier appearance-based approaches are based on gait features that can be derived from the spatio-temporal pattern of a walking person [14]. Recently, Han and Bhanu [6] proposed an energy based feature extraction technique and introduced a new concept, the gait energy image (GEI). GEI represents the temporal motion pattern within a gait cycle in a single image, that holds several key features of human gait such as motion frequency, temporal and spatial changes of the human body, as well as a global body shape statistic pattern [9]. However, the GEI, like other appearance-based algorithms, is view dependent and performs best when a side view is used [17].

There are number of approaches proposed to resolve the view dependency issues in appearance-based gait recognition. View transformation models (VTM) based on GEI proposed by Worapan *et al*. [9] adopts singular value decomposition (SVD) to transform models to different view points. However, the performance of the algorithm is good only for a limited range of view angle deviation. Another way to address the issue of view dependency is to use multiple cameras with overlapping field of views. With sufficient coverage, one could have a view which is similar enough to match with the existing closest view [10]. The performance of this algorithm also depends on how close the probe view is to an existing gallery view.

Another possible approach is performing recognition on

a reconstructed 3D model. This requires the use of multiple calibrated cameras from widely different viewpoints. The majority of 3D techniques are model-based approaches. Yamuachi *et al*. [16] used skeleton models to extract the joint angles and static parameters in 3D reconstructed data. Gu *et al*. [5] extracted the complete pose of a person in 3D using grid-based segmentation and adaptive particle filters. Working in 3D bypasses the issue of view dependency, though, the performance of these model-based methods, as well as 2D variants, are reported to be lower than recent appearance-based techniques, such as [6], using only the side view. This suggests that significant identifying information is present in the shape and the lack of appearance modelling is detrimental to the performance.

In this paper, we present an appearance-based technique, specifically, one derived from gait energy images [6]. We propose a 3D analogue of the GEI, where, instead of temporally averaging segmented silhouettes, we perform this on reconstructed voxel volumes. We term the resulting averaged volume, the **Gait Energy Volume**, or **GEV**. The recognition based on GEV features is tested on the CMU MoBo [4] database.

Having a multi-view camera setup, however, can be impractical under many applications, even in controlled conditions such as gait-based biometric authentication. An alternative to acquiring this 3D data would be to use some form of depth sensing device. We apply the GEV to partial volume reconstructions from depth images captured from a frontal viewpoint. Frontal based depth has the advantage of being able to capture essentially all characteristics of gait from a single viewpoint without the issue of self-occlusion. These depth images can be easily acquired using devices such as the Microsoft Kinect. A frontal viewpoint also makes it possible to easily integrate into biometric portals such as the one used in the multiple biometric grand challenge (MBGC) [13].

The applicability of the frontal depth based GEV is analysed using an in-house database. The CMU MoBo database is also used to simulate frontal depth GEV by synthesising appropriate volume reconstructions.

The remainder of this paper is organised as follows. Section 2 outlines the GEI algorithm, as well as our proposed GEV implementation, followed by its application on depth images in Section 3. Section 4 outlines the feature extraction and classification algorithms used in the experiments in this paper, the results of which is shown in Section 5. Section 6 concludes the paper.

2. Gait Energy Features

2.1. Gait Energy Images

Inspired by motion history images (MHI) and motion energy images (MEI) [3] proposed for action recognition, gait energy images [6] were developed for use in gait recognition. GEI represents the gait features in multiple silhouettes of a person over a gait cycle in a single image frame. The silhouettes are normalised, aligned, and temporally averaged. This forms a compact representation of a person's spatial occupancy over a gait cycle, encoding information about their gait characteristics as well as appearance, allowing identification to be performed.

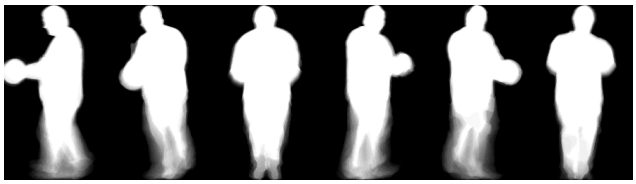


Figure 1. GEIs from multiple view angles. Higher pixel values represent greater spatial occupancy within the gait cycle.

Many extensions to the initial GEI concept have been proposed recently, including the enhanced gait energy image (EGEI) [11] and the shifted energy image (SEI) [7]. However, these algorithms are aimed to address issues such as wearing clothes [7], and do not significantly improve the performance of the underlying algorithm in situations such as view variation. Extracted GEIs from different viewpoints are shown in Figure 1.

In this paper, we use the GEI algorithm proposed by [6] to benchmark the performance of the proposed algorithms. The pixel values in the GEI are assembled into a feature vector, to which principal component analysis (PCA) and multiple discriminant analysis (MDA) are applied before classification (Section 4).

2.2. Gait Energy Volumes

The proposed gait energy volumes have many advantages over gait energy images, simply by virtue of working in 3 dimensional space. It circumvents the issue of view dependency, as well as having no pose ambiguity (e.g. left-right limbs), no self occlusion, and allowing easier segmentation of unwanted regions (e.g. hand movements). However, it is not without its disadvantages, most notably the more complex hardware setup, making it impractical for many applications.

Derived from GEIs, the construction of the GEV follows a similar process to its 2D counterpart. Binary voxel volumes, analogous to silhouettes, are spatially aligned and averaged over a gait cycle as follows,

$$\mathrm{GEV}(k) = \frac{1}{n}\sum_{t=1}^{n} \mathrm{V}(t), \tag{1}$$

where n is the number of frames in the k^{th} and V is the aligned voxel volumes. An example of binary volume and corresponding GEV is shown in Figure 2.

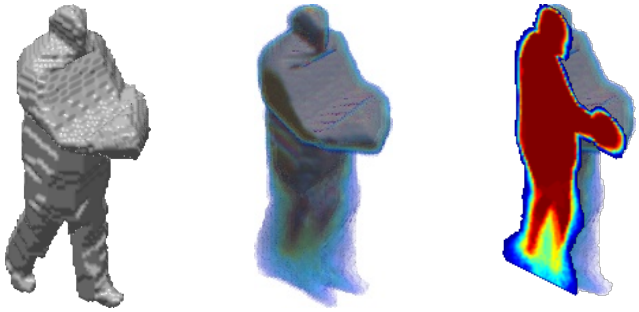


Figure 2. An example of a voxel model reconstruction (left) and GEV (middle). A cross-section representation of the GEV (right) is shown to better illustrate the internal densities.

2.3. Multi-View GEI

The proposed GEV requires multiple camera views of a subject in order to achieve a complete volume reconstruction. This gives it an advantage over the GEI in benchmarks as it has access to information not available to the GEI. In this paper, we implement a simple multi-view GEI based algorithm in order to demonstrate the advantages of extracting gait energy features in 3D.

In our algorithm, we apply the GEI to the same camera views used to construct the voxel volume used for the GEV. To combine the individual views, the feature vectors are extracted from each view's GEI and simply concatenated into a single super vector.

3. GEVs and Frontal Depth Images

Gait recognition using depth images has been attempted in the past [8]. Frontally captured data has many advantages over a lateral view for gait based biometric authentication, including easy integration into biometric portals and similar devices, as well as not having field of view issues in confined spaces such as a narrow corridor. The addition of depth also enables more data to be captured than from the side, as there is no issue of self-occlusion. In fact, all gait-based information (kinematics) can be acquired from a frontal depth sequence.

We propose applying our proposed GEV algorithm to perform gait recognition using these depth images. As only the front surface of the subject is visible, only a partial volume reconstruction is possible. The voxel model is created by taking the frontal surface reconstruction and filling to the back of the defined voxel space along the 'depth' axis. The GEV is then computed as normal using these voxel volumes.

3.1. Synthesised Depth Volumes

No gait data using depth images exists publicly. In order to test our approach, we simulate a depth reconstruction using multi-view camera data. This allows us to benchmark our results against existing gait algorithms on a common database, as well as allowing us to quantify what effects, if any, are produced by removing the back-face of the subject model.

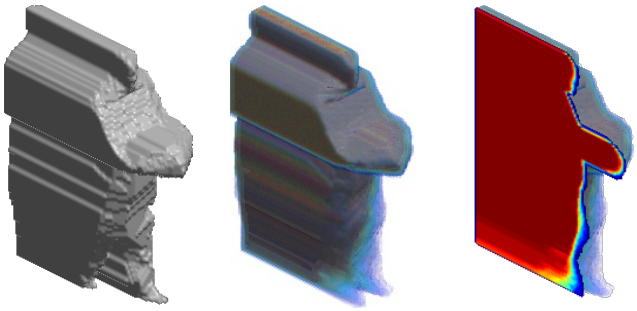
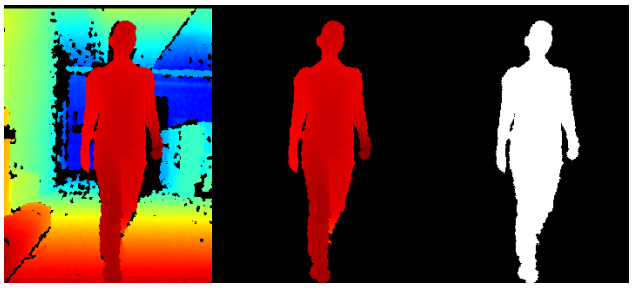


Figure 3. Synthesised depth image reconstruction and GEV.



(a) (b) (c)

Figure 4. An example of the captured database showing raw depth image (a), segmented silhouette (b), and extracted binary silhouette (c).

To generate the data used in this experiment, the front surface of the full voxel model is found (as described in Section 2.2) and filled to the back of the volume boundary. The GEV is then computed using these frontal volumes. Figure 3 shows an example of a synthesised depth volume and corresponding GEV.

3.2. Frontal Depth Gait Database

To test the application of our algorithm on real-world data, we created a small gait dataset of frontally acquired depth images[1]. The database consists of 15 subjects walking towards the camera at two different speeds, 'normal' and 'fast'. Five sequences were recorded for each subject and class, with each sequence covering an average of $2-3$ gait cycles, though only about 2 cycles is useful due to limitations in depth resolution. The dataset is captured at approximately 30 fps using the Microsoft Kinect. Colour video was also recorded but was not used. An example frame from our database is shown in Figure 4.

In order to create the GEV, we first project each pixel

[1]Contact the author at *s.sridharan@qut.edu.au* for a copy of the database.

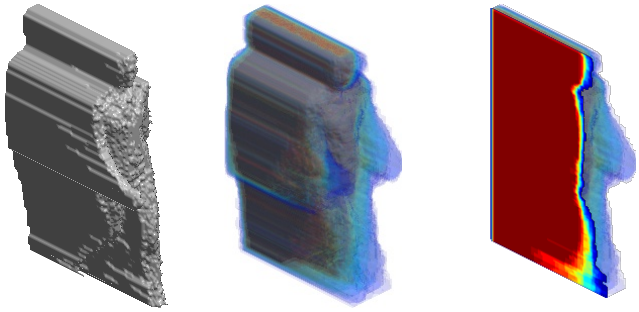


Figure 5. Volume reconstruction from depth image and GEV.

in the depth image into world coordinates, where segmentation is performed to remove the floor, ceiling and walls. From this, a surface mesh is constructed depicting the front of the subject, and alignment is performed using the centre of the torso surface. The GEV is generated as described in Section 2.2. Figure 5 shows an example of a constructed volume, as well as computed GEV.

4. Feature Modelling and Classification

4.1. Feature Modelling

Walking gait is mostly represented by the lower part of body (legs). In addition the lower part is not affected by appearance changes due to carrying goods and hand movements. Therefore, we only consider the features of the lower part of human, by considering the section below to the centroid. Extracted gait energy based features using the proposed methods are of a high dimensionality as they use image pixels or voxels. We use PCA to reduce the dimensionality, and MDA to extract the effective discriminant features as in [6].

Given n, d-dimensional (here d represents the number of image pixels for GEI based features and number of voxels for GEV based features) feature templates $X_1, X_2, ..., X_n$, PCA attempts to minimise the error function,

$$J_{d_p} = \sum_{k=1}^{n} \left\| (m + \sum_{j=1}^{d_p} a_{kj} e_j) - X_k \right\|^2, \qquad (2)$$

where d_p is the number of principal components required; m is the mean of the template features; and $e_1, e_2, ..., e_{d_p}$ is a set of orthogonal unit vectors. X_k is projected into this space to form a. The error, J_{d_p}, is minimised when $e_1, e_2, ..., e_{d_p}$ are the eigenvectors with the largest eigen values. For our proposed methods, the required dimension, d_p, is defined as number of eigen vectors required to reduce the error rate less than 1%. The original d-dimensional feature vectors are then projected to the new d_p-dimensional feature space, Y_k, using the PCA transformation matrix, T_{pca}, that contains the major eigenvector coefficients.

MDA is applied to these projected features (Y) to extract the most discriminant features to recognise the individuals. Suppose Y corresponds to c classes. MDA finds the transformation matrix, W, that maximises the ratio of between-class (inter-class) to within class (intra-class) variance. This will happen when the columns of W are the generalised eigenvectors and correspond to the largest eigenvalues of the within-class and between-class scatter matrices. The MDA transformation, T_{mda}, contains the largest coefficients of the generalised eigenvectors and projects Y to d_m-dimensional Z. The equation,

$$[Z_k]_{d_m,1} = [T_{mda}]_{d_m,d_p} \times [T_{pca}]_{d_p,d} \times [X_k]_{d,1}, \qquad (3)$$

shows the total transformation from X to Z with the fused method of PCA and MDA.

Selection of d_m depends on the number of subjects. It is obvious d_m cannot be greater than the number of classes. Therefore, d_m is set as one less than the number of classes (i.e, $c - 1$).

4.2. Classification

Since the gait energy based features represent the gait cycle, the distance, d_{ij} between i^{th} probe cycle and j^{th} gallery cycle is determined by computing the Euclidean distance between the gait cycles' feature vectors (F),

$$d_{ij} = \sqrt{\sum |F_i - F_j|^2}. \qquad (4)$$

To determine the distance between a particular probe and the gallery subject, each of which is composed of multiple gait cycles, the algorithm proposed by Boulgouris *et al.* [1] is used. From this, the distance to the closest gallery cycle from each probe cycle $\left({d_{\min}}_i^P\right)$, and the distance to the closest probe cycle from each gallery cycle $\left({d_{\min}}_j^G\right)$ are found,

$${d_{\min}}_i^P = \min_j \left(d_{ij}\right), \qquad {d_{\min}}_j^G = \min_i \left(d_{ij}\right). \qquad (5)$$

The final distance, D, between the probe and gallery sequence is,

$$D = \frac{1}{2}\left(\text{median}\left({d_{\min}}^P\right) + \text{median}\left({d_{\min}}^G\right)\right). \qquad (6)$$

5. Experiments and Results

Two different databases are used to evaluate the proposed algorithms, CMU MoBo and our in-house captured frontal depth-based gait database (see Section 3.2).

5.1. CMU MoBo Database

The CMU MoBo database [4] is used to evaluate all the systems presented in this paper. The database consists of 25

Figure 6. Examples of multi-view images from the CMU MoBo database.

Experiment	Gallery Class	Probe Class
Slow-Ball	Slow Walk	Ball
Slow-Fast	Slow Walk	Fast Walk
Ball-Fast	Ball	Fast Walk

Table 1. Inter-class test cases.

subjects under four test classes captured from 6 cameras simultaneously as they walk on an indoor treadmill. The four classes are slow walk, fast walk, walking while carrying a ball, and walking on an inclined surface, though the last is not used in this evaluation as no background image with the inclined treadmill was provided to enable clean silhouette segmentation. Examples of multi-view images from the CMU MoBo database are shown in Figure 6.

The experiments evaluate the GEVs generated from the visual hull of the multi-view silhouettes, as well as the frontal GEVs extracted through synthesised 'depth' reconstruction as described in Section 3.1. As a baseline, the GEI of the front and side views is also tested, as well as the multi-view GEI (Section 2.3).

Both intra and inter-class cases are considered and Receiver Operating Curves (ROC) are used to compare the results. In the intra-class cases, the video sequence is split in half, with the gait cycles in the first half used as the gallery and the second half used as the probe. Intra-class recognition performance of all algorithms is 100%, without any false alarms. For inter-class tests, the full sequence is used as either the gallery or probe, in the combination shown in Table 1.

ROC curves for the inter-class experiments are shown in Figure 7. From the results, it can be seen that our GEV approaches outperform the GEI. This includes the multi-view GEI, showing that it is worth while performing the simple volume reconstruction and working in a 3D space given multi-view data. The GEV applied to the synthesised depth reconstruction outperforms the full volume GEV. This is likely due to essentially all gait characteristics being acquired from a frontal perspective, while the back filling reduces the impact of noise.

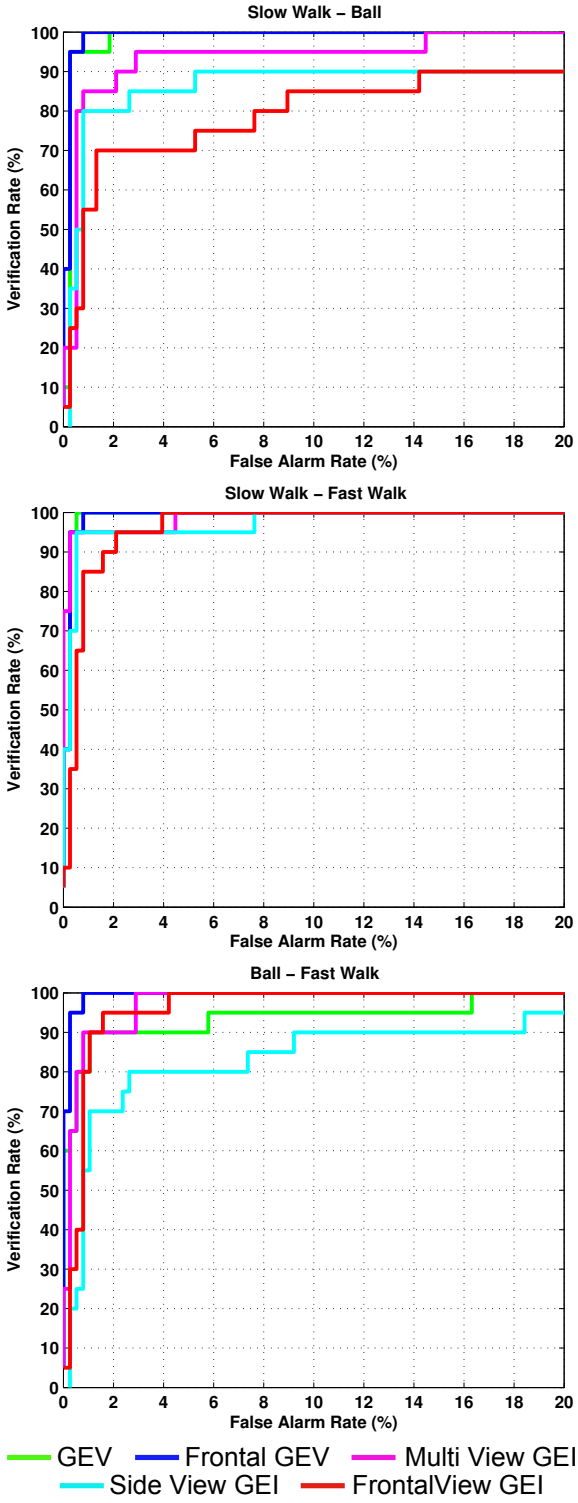


Figure 7. ROC curves for inter-class tests on the CMU Mobo database.

5.2. Frontal Depth Gait Database

Experiments evaluating the GEV on real world depth images is done using the in-house frontal depth gait database (Section 3.2), as no such data is available in the public domain. We perform tests for both inter and intra-class cases, with an inter-class experiment done with fast walk as the probe and normal as the gallery. The GEI was also tested

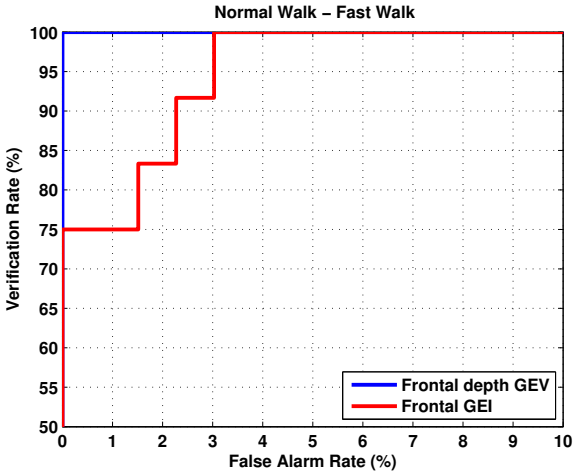


Figure 8. ROC curves for inter-class test on frontal depth gait database.

with silhouettes extracted directly from the depth images (Figure 4), as these silhouettes were of higher quality than those produced from the colour images. Due to the small database size, only a single cycle was used for both probe and gallery in order to artificially lower the performance. During intra-class tests, GEI and GEV both record 100% without any false alarms. However, the inter-class recognition rate of 100% for GEV and only 75% for GEI at false alarm rates of less than 1% (Figure 8), shows the applicability of our GEV algorithm to real world depth images.

6. Conclusion

In this paper, we propose an extension of the GEI to operate in the 3D domain, using binary voxel volumes instead of 2D silhouettes. This proposed GEV algorithm shows an improvement over its 2D variant in performing gait recognition, given multi-view data. We also demonstrated the applicability of our algorithm to frontally captured depth images, such as would be acquired using a biometric portal, through the use of synthesised data as well as recorded sequences.

We plan to expand our depth-based gait dataset to include more test subjects, as well as cover a variety of test conditions, particularly those more likely to be experienced biometric gait application, such as the carrying of luggage, changes in clothing and changes in session times. Our proposed algorithm will be re-evaluated on the expanded database.

References

[1] N. V. Boulgouris, K. N. Plataniotis, and D. Hatzinakos. Gait recognition using linear time normalization. *Pattern Recognition*, 39(5):969–979, 2006. 4

[2] 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, 2003. 1

[3] J. W. Davis. Hierarchical motion history images for recognizing human motion. In *Proc. IEEE Workshop on Detection and Recognition of Events in Video*, pages 39–46, 2001. 2

[4] R. Gross and J. Shi. The CMU motion of body (MoBo) database. Technical Report CMU-RI-TR-01-18, Robotics Institute, Pittsburgh, PA, June 2001. 2, 4

[5] J. Gu, X. Ding, S. Wang, and Y. Wu. Action and gait recognition from recovered 3-D human joints. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 40(4):1021–1033, 2010. 2

[6] J. Han and B. Bhanu. Individual recognition using gait energy image. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 28(2):316–322, 2006. 1, 2, 4

[7] X. Huang and N. V. Boulgouris. Gait recognition using linear discriminant analysis with artificial walking conditions. In *Proc. IEEE Int. Conf. on Image Processing*, pages 2461–2464, 2010. 2

[8] D. Ioannidis, D. Tzovaras, I. G. Damousis, S. Argyropoulos, and K. Moustakas. Gait recognition using compact feature extraction transforms and depth information. *IEEE Trans. on Information Forensics and Security*, 2(3):623–630, 2007. 3

[9] W. Kusakunniran, Q. Wu, H. Li, and J. Zhang. Multiple views gait recognition using view transformation model based on optimized gait energy image. In *Int. Conf. on Computer Vision Workshops*, pages 1058–1064, 2009. 1

[10] C.-S. Lee and A. Elgammal. Towards scalable view-invariant gait recognition: Multilinear analysis for gait. In *Proc. Int. Conf. on Audio and Video-Based Biometric Person Authentication*, pages 395–405, 2005. 1

[11] C. Lin and K. Wang. A behavior classification based on enhanced gait energy image. In *Proc. Int. Conf. on Networking and Digital Society*, volume 2, pages 589–592, 2010. 2

[12] M. S. Nixon, J. N. Carter, and C. Yam. Automated person recognition by walking and running via model-based approaches. *Pattern Recognition*, 37(5):1057–1072, 2004. 1

[13] P. J. Phillips, P. J. Flynn, J. R. Beveridge, W. T. Scruggs, A. J. O'toole, D. S. Bolme, K. W. Bowyer, B. A. Draper, G. H. Givens, Y. M. Lui, H. Sahibzada, J. A. Scallan, and S. Weimer. Overview of the multiple biometrics grand challenge. In *Proc. Int. Conf. on Biometrics*, pages 705–714, 2009. 2

[14] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, and K. W. Bowyer. The humanid gait challenge problem: data sets, performance, and analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(2):162–177, 2005. 1

[15] 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, 2004. 1

[16] K. Yamauchi, B. Bhanu, and H. Saito. Recognition of walking humans in 3D: Initial results. In *Computer Vision and Pattern Recognition Workshops*, pages 45–52, 2009. 2

[17] S. Yu, D. Tan, and T. Tan. Modelling the effect of view angle variation on appearance-based gait recognition. In *Proc. Asian Conf. on Computer Vision*, pages 807–816, 2006. 1