

# The University of Southampton Multi-Biometric Tunnel and introducing a novel 3D gait dataset

Richard D. Seely, Sina Samangooei, Lee Middleton, John N. Carter and Mark S. Nixon

**Abstract**—This paper presents the University of Southampton Multi-Biometric Tunnel, a constrained environment that is designed with airports and other high throughput environments in mind. It is able to acquire a variety of non-contact biometrics in a non-intrusive manner. The system uses eight synchronised IEEE1394 cameras to capture gait and additional cameras to capture images from the face and one ear, as an individual walks through the tunnel. We demonstrate that it is possible to achieve a 99.6% correct classification rate and a 4.3% equal error rate without feature selection using the gait data collected from the system; comparing well with state-of-art approaches. The tunnel acquires data automatically as a subject walks through it and is designed for the collection of very large gait datasets.

## I. INTRODUCTION

With the ever increasing demand for security and identification systems, the adoption of automatic gait recognition systems may become widespread in the future. Automatic gait recognition is attractive because it enables the identification of a subject from a distance, meaning that it will find applications in a variety of different environments. In scenarios such as access control or surveillance where there is very little control over the environments, complex computer vision algorithms are often required for analysis. However, constrained environments such as walkways in airports, where the surroundings and the path taken by individuals can be controlled provide an ideal application for such systems. Figure 1 depicts an idealised constrained environment. The path taken by the subject is restricted to a narrow path and once inside is in a volume where lighting and other conditions are controlled to facilitate gait analysis. The ability to control the surroundings and the flow of people greatly simplifies the computer vision task, compared to typical unconstrained environments.

Even though datasets with greater than one-hundred people are increasingly common, there is still very little known about the inter and intra-subject variation in gait. This information is essential to estimate the recognition capability and limits of automatic gait recognition systems. In order to accurately estimate the inter and intra class variance, substantially larger datasets are required. Covariates such as footwear type, surface type and carried items have as yet received only limited attention; although considering the potentially large impact on an individual's gait, large trials need to be conducted to establish how much variance in gait results.

School of Electronics and Computer Science, University of Southampton, UK. Contact: jnc@ecs.soton.ac.uk

As yet, there are very few demonstrable automatic gait recognition systems that can be applied to practical scenarios and performance figures for such a system are still not well known.

This paper presents the University of Southampton Multi-Biometric Tunnel; a biometric portal using automatic gait recognition for identification purposes. The tunnel provides a constrained environment and is ideal for use in high throughput security scenarios and for the collection of large datasets. A prototype design is described and early results are presented; showing the potential for the system.

## II. REVIEW

Automatic gait recognition has many unique advantages, which means that it is highly suitable for a variety of applications. An individual's gait can be measured in a non-invasive manner without explicit cooperation using standard video equipment. Gait is one of a few biometrics that can be measured from a distance; unlike many other biometrics where close contact is required. This also means that the recognition system can be mounted out of reach, making it harder to tamper with.

Medical research into gait dates back to 1964, when Murray et al. [15] used a 20 Hz strobe unit and a long exposure camera to record the trails formed by a set of marker strips attached to participant's legs, from this a basic biomechanical model was found. A more sophisticated model was later proposed by Cutting et al. [3]. Gait analysis now finds widespread use in the medical field, being used for both diagnosis and patient rehabilitation. The most common gait analysis systems in the medical field use markers attached

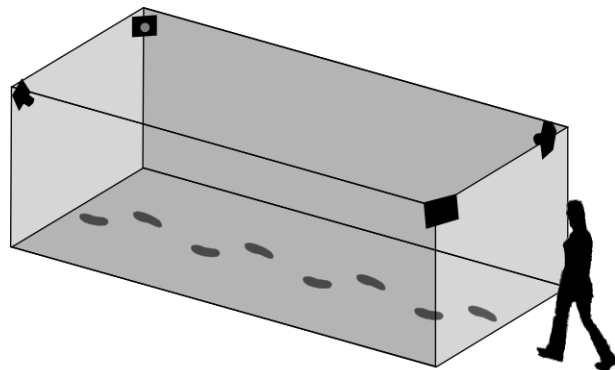


Fig. 1. A controlled environment with fixed cameras provides an ideal scenario for automatic gait recognition. The subject is constrained to walk through the middle; controlled lighting and background facilitate analysis.

to the subject, and provide three-dimensional data on the trajectories of the markers as the subject walks.

Gait has also received interest from researchers in psychology; in 1973 Johansson [6] showed that individuals could recognise human motion by observing a set of points attached to the subject on a moving light display. Soon after, it was shown that humans could distinguish the gender of the subject when observing the moving light display[7, 12].

While medical, bio-mechanical and psychology research on human gait has a long history, the use of gait as a biometric is comparatively new; one of the earliest works being that of Niyogi and Adelson [18] in 1994. Since then a variety of analysis techniques have been proposed[16, 17].

Gait analysis techniques can be broadly classified as those based on anthropomorphic models and those based on statistical models. The latter using the distribution of the silhouette and its temporal variation for analysis. One of the earliest non-model based techniques was that of Little and Boyd [10], where ellipsoids were used to describe the variation and distribution of the optical flow in the original video data of the subject. Murase and Sakai [14] applied eigenvalue decomposition to silhouettes in order to derive a new feature space (the eigen-space), and then matched subjects using their trajectory through the eigen-space over a gait cycle. Another approach is to describe the shape of the silhouette by its boundary points; Wang et al. [25] used the boundary points of the silhouettes to find the Procrustes mean shape of the silhouette sequence and then achieve recognition by comparing the Procrustes mean shapes between subjects. One of most well known gait analysis techniques is the average silhouette[11, 5, 24], silhouettes are all rescaled to a fixed size whilst retaining the aspect ratio and all translated so that the centre of mass is central. The average of the silhouettes is then calculated, and the resulting image can be treated as a feature vector and used for recognition. Due to its simplicity and remarkably good performance[24], the average silhouette is often used as a benchmark.

The first documented model-based technique was that of Niyogi and Adelson [18], where a stick model was fitted to background segmented silhouettes using a novel technique; the silhouettes were stacked to form an XYT volume and double helices were fitted each XT slices from the volume; the stick model was then fitted using the parameters from the helices. This approach added a good degree of smoothing and robustness to the analysis technique. Cunado et al. [2] proposed a simple model-based technique where the Hough transform was used to fit the model to the video data; the Hough transform provides a good level of robustness against noise and occlusion, which leads to more robust analysis. Another well known method involves fitting seven ellipses to regions of the silhouette data, then using the frequency components of the ellipsoid’s parameter variation for recognition[9].

The vast majority of gait analysis techniques (including those mentioned above) use video data captured from a single camera and as a result of this, most exhibit some degree of viewpoint dependence. This means that the direction

TABLE I  
SOME EXISTING GAIT DATASETS

Name	Subjects	Samples	Environment
MoBo[4]	25	600	Indoor, treadmill
Soton HID[23]	114	> 5000	Indoor
Gait Challenge[20]	122	1870	Outdoor
CASIA[26]	124	13640	Indoor

in which subjects walk must be carefully controlled in order to achieve acceptable recognition performance. The use of three-dimensional models and/or data can remove this dependence. Compared to two-dimensional data, it is possible to achieve a greater level of accuracy when fitting models to three-dimensional data, which allows the use of more detailed models and even biomechanical models such as that of Cutting et al. [3]. The majority of model based gait analysis techniques make assumptions about the viewpoint relative to the camera and are essentially two-dimensional.

Attempts have been made at fitting a 3D model with 33 degrees of freedom to silhouette data[1], though this is a difficult problem as it is likely that there will be many different solutions. Using silhouette data from multiple viewpoints provides a larger number of constraints, which should reduce the number of possible solutions[19]. Instead of directly applying gait analysis to the video data from multiple cameras, it is possible to reconstruct a three-dimensional approximation of the subject[8]. The resulting three-dimensional data can then be used to synthesise images from a fixed viewpoint, enabling the application of a viewpoint-dependant gait analysis technique[22]. Surprisingly little work has been done into the use of 3D data for gait analysis, this might be due to the lack of high quality 3D data available.

There are a variety of gait datasets available; some of the most notable include MoBo[4], The Gait Challenge[20], CASIA[26] and Soton HID[23], summarised in Table I. Most datasets include video data from multiple cameras; apart from the MoBo dataset, there is no synchronisation between cameras. Many of the most popular datasets were captured using camcorders and manually edited, which took a large amount of time. The MoBo dataset contains only 25 subject and 600 samples, whilst The Gait Challenge contains 122 subjects with 1870 samples, even larger still is the Soton HID dataset, containing 114 subjects and over 5000 samples. The recent CASIA dataset contains 124 subjects and 13640 samples. It has been shown that recognition rates of 100% can be achieved using the Soton HID dataset[24], excellent results have also been achieved using the Gait Challenge data[21].

All four datasets contain a limited amount of covariate data, such as footwear, surface and carried items; although there is nowhere near enough to accurately model the variance in one’s gait. In order to extend state of the art analysis methods, a larger dataset is required with a wide range of covariate data.

### III. THE MULTI-BIOMETRIC TUNNEL

#### A. Overview

The Multi-Biometric Tunnel is a unique research facility situated in the University of Southampton, it has been specifically designed as a non-contact biometric access portal[13], providing a constrained environment for people to walk through, whilst facilitating recognition. The system has been designed with airports and other high throughput environments in mind, where contact based biometrics would prove impractical. Such a system could be setup in a very unobtrusive manner where individuals might not even be aware of its presence. It also enables the automated collection of large amounts of non-contact biometric data in a fast and efficient manner, allowing very large datasets to be acquired in a significantly shorter timeframe than previously possible.

#### B. Configuration

The multi-biometric tunnel is able to detect the entry and exit of an individual, allowing a high degree of automation. Whilst a subject is inside the tunnel their gait is recorded by eight Point Grey Dragonfly cameras, allowing the reconstruction of 3D volumetric data. The gait cameras all have a resolution of  $640 \times 480$  and capture at a rate of 30 FPS, they are connected together over an IEEE1394 network employing synchronisation units to ensure accurate timing between cameras. Figure 3 shows a single frame as captured by the cameras. Video is also captured of the subject's face and upper body using a high resolution  $1600 \times 1200$  IEEE1394 camera, enabling face recognition. A  $1600 \times 1200$  snapshot is taken of the subject's side of the head, which can be used for ear biometrics. As shown in Figure 2, the facility has a central region that participants walk along, with the face and ear cameras placed at the end of the walkway and the gait video cameras positioned around the upper perimeter of the tunnel. The walls of the tunnel are painted

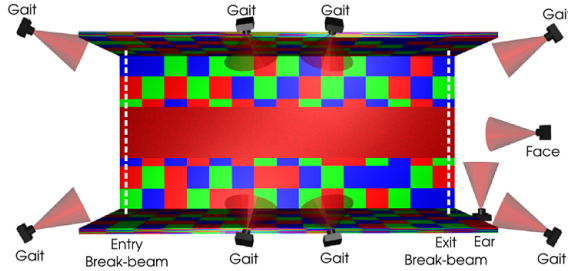


Fig. 2. Placement of cameras and break-beam sensors in system

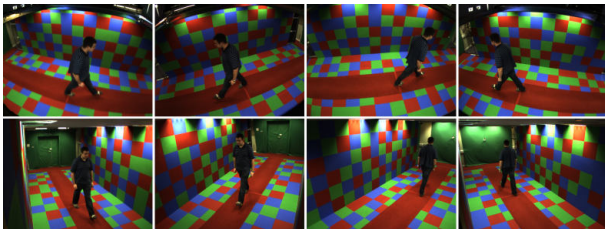


Fig. 3. Synchronised images captured by gait cameras

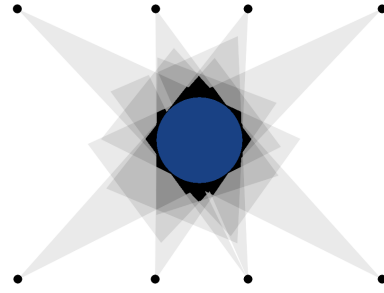


Fig. 4. Intersection of cameras

with a non-repeating rectangular pattern to aid automatic camera calibration. Saturated colours have been chosen to ease background/foreground separation. These choices are mandated by the experimental nature of the facility.

#### C. Data Processing

There are several stages to processing the video collected from the gait cameras in the tunnel. First, the raw Bayer tiled frames from the cameras are converted to colour frames using bi-linear interpolation. Separate background and foreground images are used to facilitate background subtraction and shadow suppression. This is followed by some post-processing using simple morphological operators to clean up the silhouette data. The resulting silhouettes are corrected for radial distortion and then used as the basis for shape from silhouette reconstruction[8]. Shape from silhouette reconstruction is simply the calculation of the intersection of projected silhouettes, see Figure 4, and it can be expressed mathematically as:

$$V(x, y, z) = \begin{cases} 1 & \text{if } \sum_{i=1}^N I_n(M_n(x, y, z)) \geq k \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Where  $V$  is the derived 3D volume,  $k$  is the number of cameras required for a voxel to be marked as valid and  $N$  is the total number of cameras.  $I_n$  is the silhouette image from camera  $n$  where  $I_n(u, v) \in \{0, 1\}$ , and  $M_n(x, y, z : u, v)$  is a function that maps the three-dimensional world coordinates to the coordinate system of camera  $n$ .  $M_n$  is calculated using the calibration information derived for each camera. In a conventional implementation of shape from silhouette, a voxel may only be considered valid if all cameras have silhouette pixels at its location; therefore  $k = N$  must be satisfied. However in this implementation only six out of eight cameras are required, thus  $k = 6$  and  $N = 8$ . This results in a less selective criteria, which adds a degree of resilience against background segmentation errors; although the reconstructed shape is not as accurate. A small amount of post-processing is carried out on the resulting 3D volumes using binary morphology to improve the accuracy of the reconstructed volumes. An example volume created by SFS reconstruction is shown in Figure 5.

### IV. EXPERIMENTAL SETUP

Upon arrival, the experiment's purpose and procedure was explained to each potential participant and if they agreed,

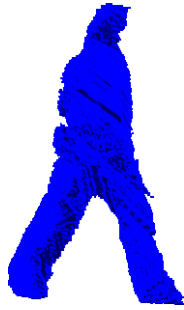


Fig. 5. A 3D volumetric frame created by the multi-biometric tunnel using SFS reconstruction

then they signed a consent form to confirm that they were willing to participate in the experiment. In order to ensure the privacy for all participants, the consent forms had no unique identifiers on them, and as such they are the only record of the participant’s identity. Each participant was asked to choose a unique identifier at random, which could then be associated with any data collected from that individual. Before commencing the experiment, each subject was asked to walk through the tunnel as a trial run, this was not recorded and was watched by the supervisor to ensure that the subject understood their instructions.

The multi-biometric tunnel was equipped with a status light mounted outside of the visible area; participants were asked to wait nearby until it indicated that the system was ready for them, upon walking through the tunnel the status light would signal busy until ready for another walk by the subject. Before each sample the gait and face cameras captured one second of video footage whilst the tunnel area was empty; this was used later for the background estimation and subtraction. Upon entering the tunnel, the subject would walk through a break-beam sensor, starting the capture process. Towards the end of the tunnel another break-beam sensor stopped the capture process. After capture, the recorded data was saved (unprocessed) to disk.

Each participant was asked to walk through the tunnel ten times, taking on average five minutes per person. After walking through the tunnel ten times, the participant was asked to answer some questions about themselves, these included questions about gender, age, ethnicity and physical parameters. The entire process of induction, walking through the tunnel and answering questions took on average 15 minutes per participant.

The experiment collected gait, face and ear data, although only the gait data is discussed here. Processing of the captured data was performed in bulk whilst the tunnel was not being used. For each sample, an automatic process was used to locate a complete gait cycle, this was achieved by analysing the variation in the size of the subject’s bounding box.

## V. GAIT ANALYSIS

Several different variants of the average silhouette gait analysis technique[11, 5, 24] were used to evaluate the

dataset collected from the multi-biometric tunnel; the normalised side-on average silhouette, the (non-normalised) side-on average silhouette and the combination of side-on, front-on and top-down average silhouettes.

The dataset used for analysis comprised of 103 subjects and contained 1030 samples, of which 1005 were valid. Of the twenty-five invalid samples, fourteen of the samples featured clipping from where the subject was outside of the reconstruction area. The other eleven samples were marked as invalid because the automatic gait cycle finder was unable to reliably identify a complete cycle. The database is made up of 73% male and 27% female subjects and the average age was 27 years.

All three gait analysis techniques discussed below have some similarities with the work of Shakhnarovich et al. [22], in that the 3D volumetric data is used to synthesise silhouettes from a fixed viewpoint relative to the subject. The resulting silhouettes are then passed to a standard 2D gait analysis technique; in this case the average silhouette. The advantage of using three-dimensional data is that silhouettes from any arbitrary viewpoint can be synthesised, even if the viewpoint is not directly seen by a camera. For example, silhouettes from an orthogonal side-on viewpoint can be synthesised from the volumetric data by:

$$J_i(y, z) = \bigcup_{x=x_{\text{MIN}}}^{x_{\text{MAX}}} V_i(x, y, z) \quad (2)$$

In other words, the side-on orthogonal viewpoint  $J_i$  for frame  $i$  is synthesised by taking the union of voxels in volume  $V_i$  along the  $x$  axis, where the  $x$  axis spans left to right,  $y$  spans front to back and  $z$  spans from the top to the bottom. In a similar manner, the front-on and top-down orthogonal viewpoints can be synthesised by taking the union of the voxels along the  $y$  or  $z$  axis respectively.

In the first analysis, silhouettes are taken from a side-on orthogonal viewpoint. This view is not seen by any camera and so must be synthesised. The use of a side-on viewpoint facilitates comparison with previous results. The average silhouette is calculated in a similar manner to that of Veres et al. [24], where the centre of mass  $\mathbf{C}_i = (C_{i,x}, C_{i,y})$  is found for each frame  $i$ . The average silhouette is then found by summing the centre of mass aligned silhouettes:

$$A(x, y) = \frac{1}{M} \sum_{i=0}^{M-1} J_i(x - C_{i,x}, y - C_{i,y}) \quad (3)$$

Where  $A$  is the average silhouette and  $M$  is the number of frames in the gait cycle. The derived average silhouette is scale normalised so that it is 64 pixels high, whilst preserving the aspect ratio. The average silhouette is treated as the feature vector and used for leave-one-out recognition, using nearest-neighbour classification and the euclidean distance as the distance metric between samples. A recognition rate of 96.6% was achieved. No feature-set transformation or selection was performed in this and subsequent analysis. This result is then similar in performance to current state-of-art approaches, yet allows other views to be analysed in future.

Because the silhouette data can be synthesised from an orthogonal viewpoint, the subject’s distance from the viewpoint will not affect the silhouette size, thus meaning that scale normalisation is unnecessary and removes valuable information. For this reason a second analysis was conducted using non scale-normalised average silhouettes, the average silhouettes were downsampled by a factor of four to reduce the computational workload. The non-normalised average silhouette retains information such as the subject’s build and height. The same viewpoint as the previous normalised variant was used, achieving an improved recognition rate of 98.6%.

The above analysis methods only utilise one viewpoint, meaning that very little of the additional information contained within the three-dimensional data was exploited. Therefore one additional analysis technique was performed, using non-normalised average silhouettes derived from three orthogonal viewpoints; side-on, front-on and top-down. The features from the three average silhouettes were simply concatenated and the resulting feature vector used for recognition, achieving an even better recognition rate of 99.6%. Again this is comparable with state-of-art approaches.

Several different analysis methods have been carried out to evaluate the quality of the collected data. The correct classification rate and equal error rate was found for each analysis method and a summary of the results is presented in Table II. The respective cumulative match scores are shown in Figure 6; it can be seen that the normalised average signature yields relatively poor performance, most likely due to the loss of information such as height and build. This is confirmed by the much improved classification performance of the non-normalised average silhouette. Classification performance using the concatenated average silhouettes proves better than both other methods, although the improvement in the equal error rate is marginal; this suggests that the additional information contained within three-dimensional data is useful for recognition.

In addition, ROC (receiver operating characteristic) curves demonstrating the system’s capability to verify identity are shown in Figure 7. These confirm that normalised side-on average silhouettes are clearly inferior. However the situation is less clear between the other two cases, where the method using multiple viewpoints proves more selective than that of a single viewpoint.

## VI. CONCLUSIONS

We have shown that the multi-biometric tunnel can be successfully used for automatic gait recognition; this demon-

TABLE II  
PERFORMANCE OF VARIOUS AVERAGE SILHOUETTE SIGNATURES ON DATASET

Average Silhouette	CCR	EER
Normalised Side-on	96.6%	9.5%
Side-on	98.6%	4.4%
[Side-on, Front-on, Top-down]	99.6%	4.3%

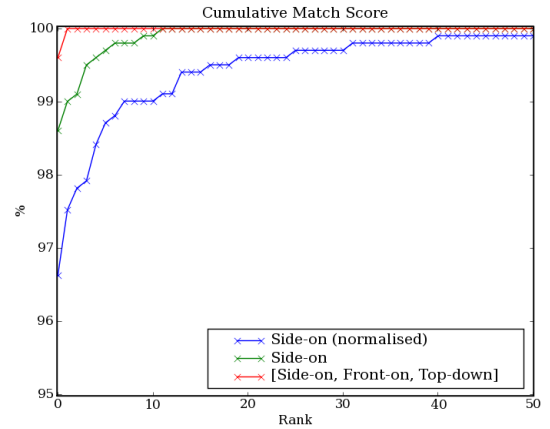


Fig. 6. Cumulative match score plots for gait analysis techniques

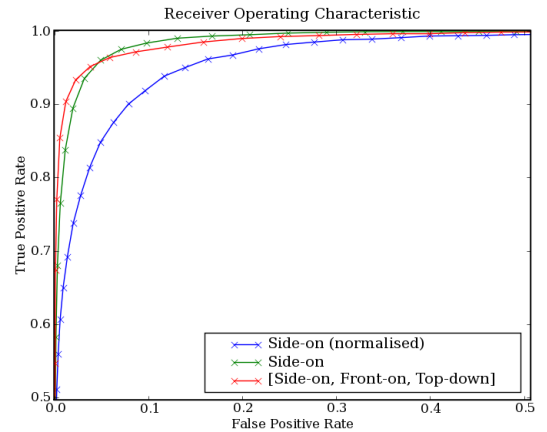


Fig. 7. Receiver operating characteristic plots for gait analysis techniques

strates its validity in non-contact biometric recognition and verification applications. By applying the simple average silhouette gait analysis technique without any feature-set selection or transformation, we achieved correct classification and equal error rates of 99.6% and 4.3% respectively. This is similar and in some cases better than other state-of-art analysis techniques on existing datasets.

We have also demonstrated that the process of scale normalisation usually employed whilst calculating average silhouettes has a negative impact on recognition performance, which is probably due to the loss of information such as the subject’s height. Also, the improvement in recognition performance achieved by using average silhouettes from multiple viewpoints proves that the additional information contained within three-dimensional data is useful.

We were able to collect a large dataset containing over 100 subjects within a period of three weeks, demonstrating that we can efficiently acquire and process gait data at a rate commensurate with collecting a very large gait database.

## VII. ACKNOWLEDGEMENTS

We gratefully acknowledge partial support by the Defence Technology Centre 8 10 supported by General Dynamics.



We would also like to thank the School of Electronics and Computer Science at the University of Southampton for two Postgraduate scholarships.

#### REFERENCES

- [1] B. Bhanu and J. Han. Human recognition on combining kinematic and stationary features. In *Proc. AVBPA*, volume 2688, pages 600–608. Springer-Verlag, 2003.
- [2] D. Cunado, M. S. Nixon, and J. N. Carter. Using gait as a biometric, via phase-weighted magnitude spectra. In *Proc AVBPA*, pages 95–102. Springer-Verlag, 1997.
- [3] J. Cutting, D. Proffitt, and L. Kozlowski. A biomechanical invariant for gait perception. *Experimental Psychology*, 4(3):357–372, 1978.
- [4] R. Gross and J. Shi. The cmu motion of body (mobo) database. Technical Report CMU-RI-TR-01-18, Robotics Institute, Carnegie Mellon University, June 2001.
- [5] J. Han and B. Bhanu. Statistical feature fusion for gait-based human recognition. In *Proc. CVPR*, volume 2, pages 842–847, June 2004.
- [6] G. Johansson. Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics*, 14:201–211, October 1973.
- [7] L. T. Kozlowski and J. E. Cutting. Recognizing the sex of a walker from a dynamic point-light display. *Perception & Psychophysics*, 21(6):575–580, June 1977.
- [8] A. Laurentini. The visual hull concept for silhouette-based image understanding. *IEEE Trans. PAMI*, 16(2): 150–162, February 1994.
- [9] L. Lee and W. E. L. Grimson. Gait analysis for recognition and classification. In *Proc. FGR*, pages 155–161. IEEE, 2002.
- [10] J. J. Little and J. E. Boyd. Describing motion for recognition. In *Proc. Computer Vision*, pages 235–240, November 1995.
- [11] Z. Liu and S. Sarkar. Simplest representation yet for gait recognition: averaged silhouette. In *Proc. Pattern Recognition*, volume 4, pages 211–214, August 2004.
- [12] G. Mather and L. Murdoch. Gender discrimination in biological motion displays based on dynamic cues. In *Biological Sciences*, volume 258, pages 273–279. The Royal Society, December 1994.
- [13] L. Middleton, D. K. Wagg, A. I. Bazin, J. N. Carter, and M. S. Nixon. Developing a non-intrusive biometric environment. In *Proc. Intelligent Robots and Systems*, pages 723–728, October 2006.
- [14] H. Murase and R. Sakai. Moving object recognition in eigenspace representation: gait analysis and lip reading. *Pattern Recognition Letters*, 17(2):155–162, February 1996.
- [15] M. P. Murray, A. B. Drought, and R. C. Kory. Walking patterns of normal men. *Bone & Joint Surgery*, 46: 335–360, 1964.
- [16] M. S. Nixon and J. N. Carter. Automatic recognition by gait. *Proc. IEEE*, 94(11):2013–2024, November 2006.
- [17] M. S. Nixon, T. Tan, and R. Chellappa. *Human Identification Based on Gait*. International Series on Biometrics. Springer, 2006. ISBN 0-387-24424-7.
- [18] S. A. Niyogi and E. H. Adelson. Analyzing and recognizing walking figures in xyt. In *Proc. CVPR*, pages 469–474, June 1994.
- [19] C. Orrite-Uruñuela, J. M. del Rincón, J. E. Herrero-Jaraba, and G. Rogez. 2d silhouette and 3d skeletal models for human detection and tracking. In *Proc. Pattern Recognition*, volume 4, pages 244–247, 2004.
- [20] P. J. Phillips, S. Sarkar, I. Robledo, P. Grother, and K. W. Bowyer. The gait identification challenge problem: data sets and baseline algorithm. In *Proc. Pattern Recognition*, volume 1, pages 385–388, 2002.
- [21] 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. PAMI*, 27(2):162–177, February 2005.
- [22] G. Shakhnarovich, L. Lee, and T. Darrell. Integrated face and gait recognition from multiple views. In *Proc. CVPR*, volume 1, pages 439–446, 2001.
- [23] J. D. Shutler, M. G. Grant, M. S. Nixon, and J. N. Carter. On a large sequence-based human gait database. In *Proc. Recent Advances in Soft Computing*, pages 66–72, 2002.
- [24] G. V. Veres, L. Gordon, J. N. Carter, and M. S. Nixon. What image information is important in silhouette-based gait recognition? In *Proc. CVPR*, volume 2, pages 776–782, 2004.
- [25] L. Wang, H. Ning, W. Hu, and T. Tan. Gait recognition based on procrustes shape analysis. In *Proc. Image Processing*, volume 3, pages 433–436, 2002.
- [26] S. Yu, D. Tan, and T. Tan. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In *Proc. Pattern Recognition*, volume 4, pages 441–444, 2006.