Performance Analysis of Gait Recognition with Large Perspective Distortion

Fatimah Abdulsattar[a,b]
[b]Al-Mustansiriya University
Baghdad , Iraq
fsa1g12@ecs.soton.ac.uk

John Carter[a]
[a]Southampton University
Southampton , UK
jnc@ecs.soton.ac.uk

Abstract

In real security scenarios, gait data may be highly distorted due to perspective effects and there may be significant change in appearance, orientation and occlusion between different measurements. To deal with this problem, a new identification technique is proposed by reconstructing 3D models of the walking subject, which are then used to identify subject images from an arbitrary camera. 3D models in one gait cycle are aligned to match silhouettes in a 2D gait cycle by estimating the positions of a 3D and 2D gait cycles in a 3D space. This allows the gait data in a gallery and probe share the same appearance, perspective and occlusion. Generic Fourier Descriptors are used as gait features. The performance is evaluated using a new collected dataset of 17 subjects walking in a narrow walkway. A Correct Classification Rate of 98.8% *is achieved. This high recognition rate has still been achieved using a modest number of features. The analysis indicate that the technique can handle truncated gait cycles of different length and is insensitive to noisy silhouettes. However, calibration errors have a negative impact upon recognition performance.*

1. Introduction

Gait is a biometric cue that can be used to identify people from remote camera according to the way we walk. Gait data can be captured in a non-invasive way without explicit intervention. Hence, it is highly suitable for surveillance and access control scenarios. In these scenarios, the subject can be recorded by cameras with different settings. This causes changes in the appearance of the subject, which in turn affect the performance of gait recognition techniques that are designed to work on a particular view. Furthermore, similar postures of the same subject will appear to be different if recorded at different positions relative to the camera. Most of the techniques proposed to solve the problem of view variation assume the appearance and orientation of the subject remains unchanged (weak perspective) within one gait cycle. However, this is often not true in real world data.

To further understand this problem, a preliminary study was done by the authors where a subset of 43 subjects walking along straight line from the Southampton Multi-Biometric Dataset [12] was used for the analysis. The silhouettes were captured by a set of 4 wide-angle cameras positioned close to the walking path and a set of 8 small-angle cameras further away from the walking path. In order to evaluate the performance for each of the 12 views, a 3D model was reconstructed using all the views except the target view which is then projected into a target view to generate a synthetic silhouette. Later, the Gait Energy Image, GEI, is extracted from the synthetic silhouettes and subject silhouettes for comparison. The average recognition rate was 42% from the wide-angle cameras set (having a strong perspective effect) and 97% from the small-angle cameras set (exhibiting weak perspective). These results reveal that the traditional technique (e.g. GEI) cannot deal with the silhouettes of high perspective distortion. Fig. 1 shows example for the silhouettes captured by one of the wide-angle cameras with their GEI. As it can be seen there is a transition in the local observation view within one gait cycle. The variation in perspective, appearance and occlusion is also prominent. This is due to the pose of the camera and the distance between the camera and the subject, which is also confirmed in [1].

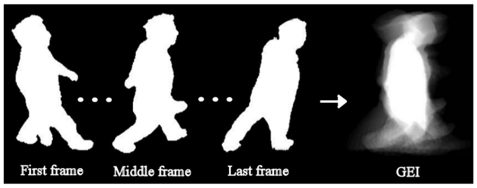


Figure 1: several frames within a gait cycle with their GEI

In order to solve this problem, we propose a gait recognition technique to identify subjects captured by an arbitrary perspective camera mounted on the wall in a constrained narrow corridor using a multiple of 3D models. The main contribution of this study is the elimination of the effect of appearance changes due to perspective and pose changes when considering gait data captured by cameras of different settings. To this end, a 3D alignment strategy is presented to match 3D models with 2D silhouette images in one gait cycle by estimating the amount of displacement between a 3D and 2D gait cycles in a 3D space. The gait features are extracted on a silhouette basis using Generic Fourier Descriptors. To test this, a 3D-2D gait dataset has been recorded which involves 3D volumetric sequences and highly perspective distorted silhouettes from an arbitrary camera. Finally, extensive analyses have been conducted to demonstrate the effectiveness of matching 3D volumes with perspective distorted silhouettes.

2. Related works

This section reviews some of the proposed techniques that tackle the problem of view variation. These techniques can be grouped into three main categories. The techniques in the first category aim to extract view-invariant gait features. Goffredo et al. [2] extracted lower limbs' poses using marker less motion estimation, which were then reconstructed in the sagittal plane using viewpoint rectification. The main limitation of this method is that the limbs' poses is untraceable in frontal view. Jeong et al. [5] used planar homography to reduce the dependency of walking direction with respect to the camera optical axis. When the walking direction is nearly parallel to the optical axis, this technique seems not to be applicable.

The second category depends on learning relationship of gait features under different views. In [8, 6] a discrete View Transform Model (dVTM) was used to transform gait features under various views into a same view. Hu [3] derived uncorrelated discriminative features across different views from tensorial data by solving tensor to vector projection via iterative processes. The techniques proposed in the second category can better cope with the problem of view variation than these in the first category, however their performance degrades as the view variation increases. Moreover, these techniques depend mainly on learning relationship between views therefore it might be difficult to recognise untrained views. Recently, Muramatsu et al. [9] built Arbitrary VTM's (AVTM) based on 3D models and a part-dependent view selection scheme to recognise untrained (arbitrary) view. This technique can perform well for cross-view matching in many settings, although its performance is degraded when the view difference is large.

The techniques in the third category gather information from multiple cameras to reconstruct a 3D human model, from which the silhouettes under any view can be synthesised. Iwashita et al. [4] proposed adaptive virtual image synthesis to identify people walking along straight and curved trajectories. Affine Moment Invariants were computed as gait features. The advantage of this method is that it used a multiple of 3D human models for training and the 2D silhouettes for testing. However, the list of test views are used to build 3D models for training and the 2D silhouettes used for testing do not have high perspective distortion. Recently, Lpez-Fernndez et al. [7] proposed a method to align 3D volumes of people walking along unconstrained path in indoor environments. The 'Gait entropy volume' is extracted as a new gait descriptor and identified using a support vector machine. Although, this method is able to cope with the observation view changes, it can be applied only on the 3D volumetric data which are not always available in real world situations. Although the techniques in the third category require a number of synchronised and calibrated cameras to build 3D human model, they can deal with a full range of views since synthetic silhouette under any view can be built to match the probe view silhouette.

Alternatively, point cloud data captured by depth sensor devices can also be used to build 2.5D model. Sivapalan, S. et al. [13] computed the Gait Energy Volume based on partial volume reconstructions created from frontal depth images. Nakajima, H. et al. [10] proposed a novel gait feature from depth images that are aligned according to virtual viewpoints to cope with view variation and perspective distortion problem. However, depth-based sensor approaches assume both gallery and probe are captured by depth sensor and deal with limited range of views.

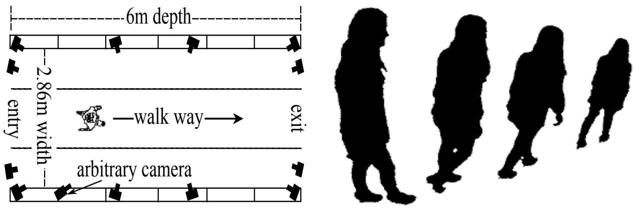
3. A new 3D-2D gait dataset [1]

Currently, the gait dataset collected by Iwashita et al. [4] contains 3D and 2D gait sequences captured by 16 synchronised cameras. However, the silhouettes from all 16 views are used to build 3D data and they do not have high perspective distortion. This means that there is no independent camera to provide perspective silhouettes from a different view. Therefore, a new data set was collected that include 3D data and 2D perspective silhouettes from an additional camera. To simulate an access control scenario in a narrow corridor (e.g. airport), a long corridor with a narrow walkway was used as a recording site, whose height, width and depth are $2.4\,\mathrm{m}$, $2.86\,\mathrm{m}$ and $6\,\mathrm{m}$ respectively. Twelve synchronised cameras were used to record each subject's gait to build 3D volumes, which have a resolution of 640×480 pixels and a frame rate of 30 frames/second. An arbitrary camera with a wide-angle was used to capture the subject's gait from behind as shown in Fig. 2(a). This camera is mounted at a height of $1.75\,\mathrm{m}$ and at a distance of $1.43\,\mathrm{m}$ from the walking path. It had a resolution of 1024×768 pixels, a frame rate of 30 frames/second and was not synchronised with the other cameras. Fig. 2(b) shows several silhouettes of the same subject captured by this camera where the effect of perspective distortion is prominent.

The new (3D-2D) gait dataset includes 17 subjects (4 women and 13 men). The subjects participated in the experi-

[1] contact the author to obtain a copy

ment with their normal clothes and were asked to walk along a straight path. Each participant walked ten times from the entrance to the exit of the studio. This provides a total of 170 multi-view gait sequences for a gallery and 170 single-view gait sequences for a probe.



(a) placement of arbitrary camera (b) examples of silhouettes

Figure 2: Recording site

4. The proposed identification technique

The proposed technique depends on a 3D alignment and projection step to allow synthetic silhouettes and subject silhouettes in one gait cycle to share the same perspective, appearance and occlusion. This is done by estimating the position of each pair of 3D and 2D gait cycles in a 3D space to determine the amount of displacement between them. All 3D volumes in one gait cycle are then aligned and projected into a 2D space of the probe camera. After that, gait features are extracted from the synthetic silhouettes and subject silhouettes using Generic Fourier Descriptors and compared using Dynamic Time Warping. Finally, the subject is recognised based on minimum distance between gait features. The details of the key operations will be explained in the following subsections.

4.1. 3D alignment and projection

The aim of 3D alignment and projection step is to make a 3D volumetric cycle at the same distance from the camera as the 2D cycle so that the virtual silhouettes share the same appearance with the real silhouette. The displacement between the 3D and 2D cycle is determined by estimating the position of the first frame in each of these cycles in a 3D space.

Here, we assume that the subject is walking along a straight line which is parallel to the z-axis as shown in Fig. 3, and the probe camera has been calibrated so that the camera's position (C), rotation (R) and internal parameters (A) are known. The 2D and 3D gait cycles are constrained so that their phases matches (i.e. the first and last frames have the same phase).

The position of a 3D gait cycle is determined by calculating the centroid of the first volume in the cycle, denoted (x_c, y_c, z_c) where x-axis spans from right to left, y-axis spans from bottom to top and z-axis spans along the direction of walking.

To estimate the position of a 2D gait cycle along the z-axis, the centroid of the first silhouette (u_c, v_c) is first computed and then projected onto a plane $x = x_c$ [2] using

$$\hat{z_c} = F(u_c, v_c; x_c, P) \tag{1}$$

where $F(.,.; x_c, P)$ is the function that projects the centroid (u_c, v_c) onto the plane $x = x_c$ using camera projection matrix P. This function can be derived from the general perspective projection equation as follows

$$\begin{aligned} \lambda\, U &= AR[I \mid -C]\, X \\ X &= \lambda R^{-1} A^{-1} U + C \end{aligned} \tag{2}$$

where λ is positive scaling factor, while $U = (u, v, 1)$ and $X = (x, y, z, 1)$ are the 2D and 3D points in homogeneous coordinates. After that, the amount of displacement (Δz) (in voxels) between 2D and 3D gait cycle along z-axis can be determined as

$$\Delta z = \hat{z}_c - z_c \tag{3}$$

To synthesise a virtual silhouette, each 3D volume V_k (k is the index of the frame in a 3D gait cycle) over one gait cycle is translated along the z-axis by Δz

$$\grave{V_k}(x, y, z) = V_k(x, y, z + \Delta z) \tag{4}$$

and then projected onto a 2D image space of the probe camera using its projection matrix. After that the frame-by-frame comparison is performed.

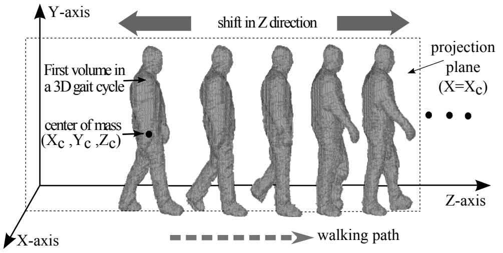


Figure 3: Shift a 3D gait cycle along z-direction

4.2. Feature extraction and comparison

The gait features are extracted from each frame individually since the pose of synthetic and real silhouettes varies from frame to frame in one gait cycle. This is in contrast to computing a global statistic (e.g. GEI). We computed Generic Fourier Descriptors (GFDs) from each silhouette over one gait cycle as gait features since GFDs are invariant to many geometric distortions (e.g. translation, scale and rotation) including a small amount of perspective, and have some intrinsic resistance to noise [14]. The GFDs can capture multi-resolution fine features in both radial and circular directions, and require a small number of features to efficiently describe the shape. To derive the GFD, first

[2] $x = x_c$ is the middle of the walking path

the image $f[x, y]$ is represented in polar coordinates $F[r, \theta]$ where $[x, y] \equiv [r \sin\theta, r \cos\theta]$, r and θ are the polar coordinates and x and y are the Cartesian coordinates of the image. To account for translation invariance, the centroid of the silhouette is set as the origin of the polar space. The Polar Fourier Transform (P) is applied as

$$P(\rho, \phi) = \sum_{r=0}^{R-1} \sum_{i=0}^{T-1} F(r, \theta_i) \times \exp\left[-j2\pi\left(\frac{r}{R}\rho + \frac{2\pi i}{T}\phi\right)\right] \tag{5}$$

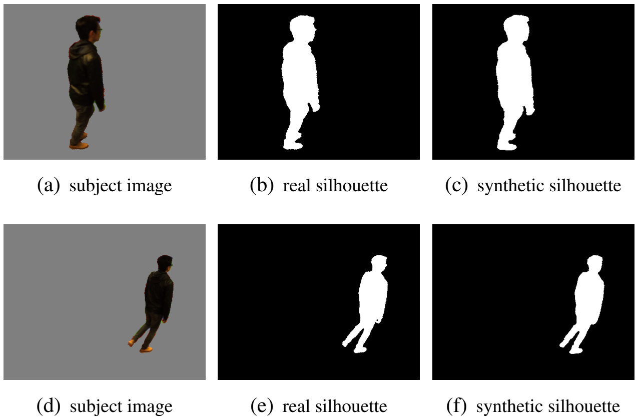
where $\theta_i = i(2\pi/T)$, $0 \leq \rho \leq R$, $0 \leq \phi \leq T$. R and T are the radial and angular resolutions respectively. To make the transform coefficients invariant to rotation and scale, the magnitude of the coefficients is calculated such that the magnitude of the first coefficient is normalised by the polar image area and the magnitudes of the remaining coefficients are normalised by the magnitude of the first coefficient. To improve the computational speed, All the original and synthetic silhouettes are centered and cropped to a fixed size. Then, only the first 4 radial and 15 angular frequencies are computed from each silhouette as gait features. As the number of frames in each gait cycle may differ due to the speed of walking, the resulting feature vectors from each gait cycle may have different lengths. The Euclidean distance and Dynamic Time Warping (DTW) [11] are used to find the minimum accumulative distance between each pair of gait features vectors.

The main differences of our work from that described in [4] is that they extracted the foot positions from each frame in a 2D and 3D cycle. Then, they used these iteratively to minimise the difference between the synthetic and actual silhouettes.

5. Results

This section presents performance analysis for the proposed technique. A Nearest Neighbour classifier was used to find the recognition results. To identify each probe sequence, we removed the corresponding 3D volumetric sequence that was captured at the same time from the gallery. Fig. 4 shows actual images (a,d) captured by the arbitrary camera within one gait cycle along with their synthetic silhouettes (c,f). There is an obvious transition in local observation view from approximately side-view to rear-view between the first and the last image. As it can be seen, the appearance and position of the synthetic silhouettes by the proposed method is similar to that of the subject images. The Correct Classification Rate (CCR) of the proposed method is 98.8%. There are only two samples misclassified by the proposed method. A visual inspection of the two misclassified samples reveals that the subjects in these samples did not walk on the straight line and thus did not obey the rules of the experiment. This leads to an error when computing the displacement for the 3D volumetric data and subsequently creates a discrepancy between the synthetic and the original silhouettes. In comparison, the GEI only obtains about 80% with this gait data.

Finally, the key and time consuming part of this technique is the projection from 3D to 2D space of the model, taking about 0.2 of the second per projection. This can be speeded up by exploiting the parallel processing capability of modern graphics cards. In the following sub-sections, we described a set of experiments to analyse the performance under various conditions.


(a) subject image (b) real silhouette (c) synthetic silhouette

(d) subject image (e) real silhouette (f) synthetic silhouette

Figure 4: Examples of real images and synthetic images by proposed method in one gait cycle.

5.1. Number of features versus performance

It is critical to select an optimal number of GFD parameters. To that end we have studied the Correct Classification Rate rate over a range of different numbers of radial and angular features. The findings are listed in Table 1 where it can be seen that high CCR's rates can be achieved using many different combinations of features. Even with a low number of features, 18 per frame, the recognition rate only drops to 93%. Table 1 shows that while the performance is improved by increasing the number of angular features (T); there is no improvement when the number of angular features exceeds 15. Meanwhile, the performance is not enhanced by increasing the number of radial features (R) for constant angular resolution. This suggests that angular features have more influence on recognition performance than radial features. These results indicate that a suitable resolution to achieve effective recognition performance with the smallest number of features is obtained by using 4 radial features and 15 angular features.

5.2. The effect of noisy silhouettes

The measurements were made in a constrained environment with good lighting and digital recording. Therefore, it is important to investigate the sensitivity of the technique to noisy silhouettes. An experiment was conducted to test

Table 1: CCR and the number of misclassified between brackets against different GFD resolutions

		R		
		3	**4**	**5**
T	**6**	93.5(9)	93.5(9)	93.5(9)
	8	95.0(7)	95.0(7)	95.0(7)
	10	96.4(5)	96.4(5)	95.7(6)
	12	96.4(5)	96.4(5)	95.7(6)
	15	97.8(3)	**98.8(2)**	97.8(3)
	17	95.0(7)	**98.8(2)**	97.8(3)

this where we added salt and pepper noise at different rates (from 10% to 40%) to the probe silhouettes, while the synthetic silhouettes in the gallery remained unchanged. Each experiment was repeated three times. For each noise rate, the averaged recognition rate and standard deviation were calculated as shown in fig. 5. As can be seen, the performance degrades slightly as the noise rate increases. An averaged CCR ranges between 96.8% and 97.8%. The maximum degradation in performance is only about 2% for 40% added noise. These findings indicate that the technique is insensitive to this type of noise. This may perhaps due to consider the whole shape and the implementation of Fourier Transform in feature extraction process.

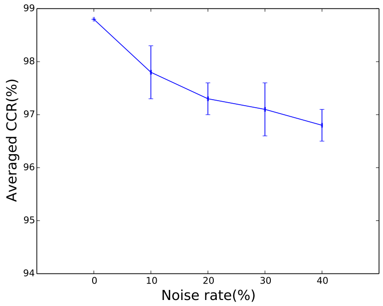


Figure 5: The impact of noisy silhouettes.

5.3. The effect of truncated gait cycle

There is a little information in the literature about how the recognition performance can be influenced by reducing the number of frames in one gait cycle and which part of the cycle contains more discriminative information. The influence of using a truncated gait cycle has been measured and the recognition performance in different parts of a gait cycle has been evaluated. To do that, we considered proportions of 25%, 50% and 75% of the total frames from each gait cycle. These proportions were selected from the start, middle and end of each gait cycle to evaluate the effective part of a gait cycle for gait recognition. The results are illustrated in fig. 6. As it can be seen that using a truncated gait cycle slightly degrades the performance where the maximum drop in recognition rate is about 3% when only 25% of the frames in a gait cycle are used. In general, the performance is slightly improved by increasing the number of frames as there is more information available for recognition. However, there is a fluctuation in performance for the frames at the middle of a gait cycle where the highest recognition rate is recorded when using only 25% of the frames and the lowest is achieved when using 50%. This fluctuation might be attributed to the small number of samples in the collected data.

The average CCRs over different proportions at the start, middle and end of a gait cycle are 96.6%, 97.1% and 95.9% respectively. The performance at the end of a gait cycle is lower than that from other parts. This is probably due to the smaller resolution of the silhouettes at this part (the subjects walk away from the camera) which could result in losing fine details and producing less discriminative features. On the other hand, the best recognition performance is achieved at the middle of a gait cycle. By observing the middle part in all gait cycles, it was found that this part includes the unoccluded motion of the leg compared to the other parts which may lead to extract more discriminative features. These results demonstrate the efficiency of the proposed technique in handling a truncated gait cycle.

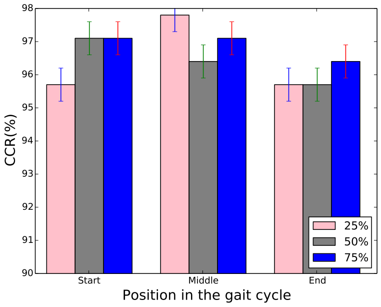


Figure 6: Performance vs. truncated gait cycle.

5.4. The Effect of camera calibration errors

In order to effectively match the synthetic silhouettes with the original silhouettes from an arbitrary camera, the synthetic silhouettes should be sufficiently accurate. This probably depends on the quality of calibration for the arbitrary camera at the first place, which in turn affects on the recognition performance. Therefore, an analysis had been performed to investigate the sensitivity of recognition performance towards calibration errors from the arbitrary camera. A translational error was introduced to the calibration matrix, and then the gait data in the gallery and probe were processed accordingly. The original calibration points, which include 2D-3D corresponding points, were used to interpret the change to the calibration matrix. The 3D points were projected into the image plane using the erroneous calibration matrix and the mean distance error between the original 2D points and the projected 2D points were calculated for different error levels.

Fig. 7 shows how the performance significantly degrades as the amount of translational error increases. When the mean distance error is less than 6 pixels, the recognition rate does not fall below 90%. The built in invariance of the chosen features compensate for small transformation errors in the projection process. However, there is a sharp decrease in recognition rate for higher error values to 78% and then to 53%. These findings reveal the sensitivity of the technique towards calibration errors, which emphasise the importance of accurately characterising the camera. This is expected as the matching step mainly depends on the accuracy of 3D alignment and projection step. Hence, the presence of error in the calibration matrix could affect on the amount of displacement between 2D and 3D gait cycles, which in turn creates discrepancy between the synthetic and the original silhouettes.

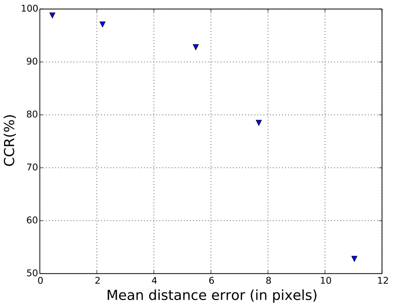


Figure 7: The effect of calibration error upon performance

6. Conclusions

We have been proposed a technique for gait recognition using a wide-angle arbitrary perspective camera in a constrained narrow walkway. A 3D alignment and projection algorithm was used to remove the variation in shape, perspective and occlusion. The performance was evaluated using a new gait dataset, which showed a high recognition capability. We observed that this algorithm depends on people walking on a straight path. The analysis illustrate that higher performance can be achieved using modest number of features. The technique can also efficiently handle truncated walking cycles and it is noise reliance. However, the calibration errors have a negative impact upon performance. In future, an improved technique to tackle the problem of direction changes in one gait cycle will be developed.

References

[1] N. Akae, Y. Makihara, and Y. Yagi. The optimal camera arrangement by a performance model for gait recognition. In *IEEE International Conference on Automatic Face Gesture Recognition and Workshops*, pages 292–297, March 2011.

[2] M. Goffredo, I. Bouchrika, J. N. Carter, and M. S. Nixon. Self-calibrating view-invariant gait biometrics. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 40(4):997–1008, 2010.

[3] H. Hu. Multiview gait recognition based on patch distribution features and uncorrelated multilinear sparse local discriminant canonical correlation analysis. *IEEE Transactions on Circuits and Systems for Video Technology*, 24(4):617–630, April 2014.

[4] Y. Iwashita, K. Ogawara, and R. Kurazume. Identification of people walking along curved trajectories. *Pattern Recognition Letters*, 48:60–69, 2014.

[5] S. Jeong, T. Kim, and J. Cho. Gait recognition using description of shape synthesized by planar homography. *The Journal of Supercomputing*, 65(1):122–135, 2013.

[6] W. Kusakunniran, Q. Wu, H. Li, and J. Zhang. Multiple views gait recognition using view transformation model based on optimized gait energy image. In *IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops)*, pages 1058–1064, 2009.

[7] D. Lpez-Fernndez, F. Madrid-Cuevas, A. Carmona-Poyato, R. Muoz-Salinas, and R. Medina-Carnicer. Entropy volumes for viewpoint-independent gait recognition. *Machine Vision and Applications*, pages 1–16, 2015.

[8] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, and Y. Yagi. Gait recognition using a view transformation model in the frequency domain. In *Computer Vision-ECCV 2006*, pages 151–163. Springer, 2006.

[9] D. Muramatsu, A. Shiraishi, Y. Makihara, M. Uddin, and Y. Yagi. Gait-based person recognition using arbitrary view transformation model. *IEEE Transactions on Image Processing*, 24(1):140–154, January 2015.

[10] H. Nakajima, I. Mitsugami, and Y. Yagi. Depth-based gait feature representation. *IPSJ Trans. on Computer Vision and Applications*, 5:94–98, July 2013.

[11] S. Salvador and P. Chan. Toward accurate dynamic time warping in linear time and space. *Intelligent Data Analysis*, 11(5):561–580, October 2007.

[12] R. D. Seely, S. Samangooei, M. Lee, J. N. Carter, and M. S. Nixon. The university of southampton multi-biometric tunnel and introducing a novel 3d gait dataset. In *2nd IEEE International Conference on Biometrics: Theory, Applications and Systems, BTAS*, pages 1–6, 2008.

[13] S. Sivapalan, D. Chen, S. Denman, S. Sridharan, and C. Clinton Fookes. Gait energy volumes and frontal gait recognition using depth images. In *1st IEEE International Joint Conference on Biometrics*, pages 1–6, Washington DC, USA, October 2011.

[14] D. Zhang and G. Lu. Shape-based image retrieval using generic fourier descriptor. *Signal Processing: Image Communication*, 17(10):825 – 848, 2002.