3D Ellipsoid Fitting for Multi-view Gait Recognition

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 recognition approaches continue to struggle with challenges including view-invariance, low-resolution data, robustness to unconstrained environments, and fluctuating gait patterns due to subjects carrying goods or wearing different clothes. Although computationally expensive, model based techniques offer promise over appearance based techniques for these challenges as they gather gait features and interpret gait dynamics in skeleton form. In this paper, we propose a fast 3D ellipsoidal-based gait recognition algorithm using a 3D voxel model derived from multi-view silhouette images. This approach directly solves the limitations of view dependency and self-occlusion in existing ellipse fitting model-based approaches. Voxel models are segmented into four components (left and right legs, above and below the knee), and ellipsoids are fitted to each region using eigenvalue decomposition. Features derived from the ellipsoid parameters are modeled using a Fourier representation to retain the temporal dynamic pattern for classification. We demonstrate the proposed approach using the CMU MoBo database and show that an improvement of 15-20% can be achieved over a 2D ellipse fitting baseline.

1. Introduction

Recognising people is a challenging task in computer vision. 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. Unlike other biometrics, gait also has the potential to be used in a surveillance environment with low-resolution video [6].

Gait can be defined as a coordinated, cyclic combination of movements that results in human locomotion [3]. The movements are coordinated in the sense that they must occur with a specific temporal pattern for the gait to occur. The walking gait cycle of a particular leg describes the movements that takes place during the walk, from the time of one heel touching the ground until the same heel retouches the ground as illustrated in the Figure 1.

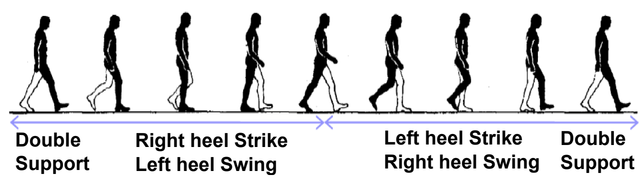


Figure 1. The Gait Cycle[1].

There are two major approaches to gait recognition; appearance based (model free) and model based [14]. The majority of early approaches were appearance based, using features such as the silhouette. These approaches are generally less complex, allowing faster processing. However, as they do not gather gait dynamics directly, they are more sensitive to background noise and are view dependent. Model based approaches are less sensitive to the above-mentioned conditions, but are generally much more complex and computationally demanding.

Current state of the art gait recognition techniques can achieve upwards of a 90% recognition rate under controlled conditions with a small number of subjects. However, under unconstrained environments with changes in illumination and in outdoor settings the recognition rate falls off rapidly. Since appearance based techniques cannot handle these variation effectively, recent research on gait recognition has focused predominantly on model-based approaches and inroads have been made in this direction. However, there is significant room for further improvement as robustness to all variations has not been achieved.

To address some of the challenges of unconstrained gait recognition, we propose a multi-view gait recognition technique based on a novel 3D ellipsoid fitting algorithm, which

[1] Reproduced from
http://www.laboratorium.dist.unige.it/ piero/Teaching/Gait

uses ellipsoidal parameters as gait features. This approach extends the work of Lee and Grimson [13], who proposed an approach to fit ellipses to silhouettes. This and other recent extensions [10], however, all operate within the 2D domain, are limited to orthogonal views only, and thus are heavily view dependent.

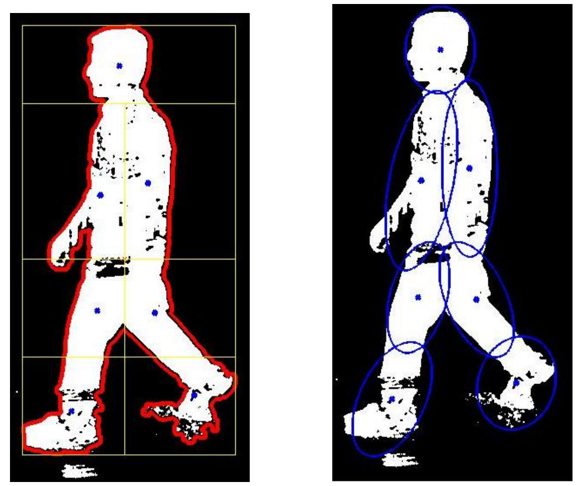
In this paper, we enable these ellipsoidal techniques to operate in the 3D domain to directly overcome both limitations of view dependency and also self-occlusion. In our proposed approach, a voxel model is constructed from the visual hull created by silhouettes from multiple views. The lower limbs are segmented and ellipsoids are fitted to these regions. The ellipsoid parameters for each limb segment are extracted and form the features used in the gait recognition process. The use of this 3D model approach is superior as it allows individual gait cycles of the left and right strides to be detected, segmented and modelled. The approach also significantly improves performance when there is a class mismatch between the gallery and probe sequences. Finally, Fourier harmonics of gait features are computed to compare subjects. Similarity values for classification are calculated from Euclidean distance between the complex harmonic values.

The proposed algorithm presented in this paper, addresses the limitations in the model based and appearance based techniques by combining the strengths of each approach. The fitting of ellipsoids to a voxel model performs much faster than full pose estimation systems, yet still provides a direct estimate to the underlying kinematic features (joint angles). The move to 3D space solves the issue of view dependency and problems with self occlusions, though our method does constrain its applications to situations where a multi-camera setup is in existance. While we demonstrate the proposed ellipsoid fitting approach to perform gait recognition, it should be noted that the model could also be used in other applications such as gesture recognition.

The remainder of this paper is organised as follows. In Section 2, we outline existing gait recognition techniques. 3D human model reconstruction and segmentation is described in Section 3. Section 4 describes the gait feature extraction process and gait period detection. Gait recognition and classification and experimental results are followed in Section 5 and Section 6. Section 7 concludes the paper.

2. Related Work

In this section we provide a brief overview of the two main approaches to gait recognition: appearance based and model based. A variety of appearance based approaches have been proposed, many of which use silhouette images as the feature. Kale *et al*. [12] used a Hidden Markov Model (HMM) to model gait features and defined the concept of Frame to Exemplar Distance (FED). BenAbdelkader *et al*. [1] proposed the Self Similarity Plot (SSP) to encode the projection of gait dynamics. The feature vectors that were used for classification consisted of units of self-similarity of size one period, and probe sequences are scaled to a constant length to compensate for period differences.



(a) (b)

Figure 2. Ellipse fitting to silhouettes.

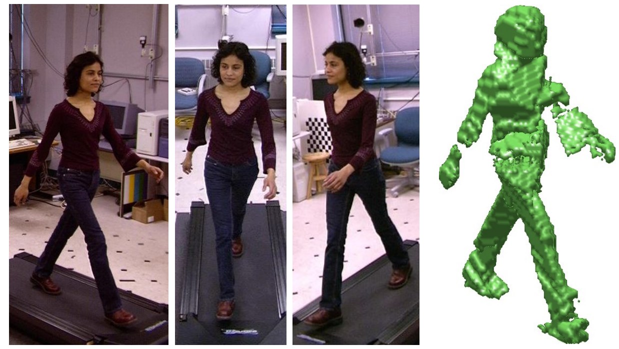
Recent appearance based approaches have used the Gait Energy Image (GEI) [9]. The GEI is the average silhouette taken over a single gait period, enabling the temporal information of gait to be encoded in a single frame. As such, it is less sensitive to silhouette noise in individual frames and is less affected by varying gait periods. Various modifications to GEI have been proposed [17, 18], however these, as well as all other appearance based techniques, suffer from view dependency.

Model based techniques are based on structural and dynamic parameters of human gait. These techniques aim to represent the human body parts - the head, torso, hip, thigh, knee and ankle - with primitive shapes (cylinders, cones, and blobs) and measurements of length, width and position. Motion models can also be incorporated to describe the kinematics of the motion of each body part, enabling more accurate estimation of the pose to be obtained.

Cunado *et al*. [4] proposed a motion-based model to analyse the angular motion of the hip and thigh by means of a Fourier series. Wagg and Nixon [16] proposed bulk motion and shape estimation guided by biomechanical analysis and used mean gait data to create the motion models. Guoying *et al*. [8] used skeleton models to extract the joint angles and static parameters. They defined the model with a finite number of degrees of freedom that can be extracted from multi-view 2D images. Junxia *et al*. [11] extracted the complete pose of a person in 3D for action and gait recognition using grid based segmentation and adaptive particle filters.

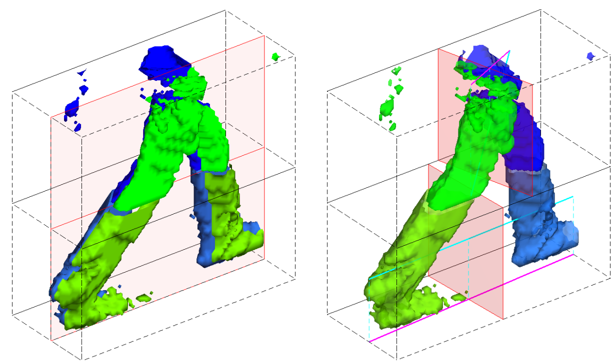
Lee and Grimson [13] proposed fitting ellipses to the silhouette pixels and extracted moment based region features. Rather than taking the entire silhouettes, they divided the silhouette into regions and fitted ellipses based on the statistics of the region. Major and minor axes of the ellipses were defined using eigenvectors and eigenvalues based on the covariance matrix of each region. Division of regions and the ellipse fitting is illustrated in Figure 2. Gait recognition is performed using the centroid, angle and axis length ratios of the ellipses. It is shown that these simple features are enough to discriminate between subjects based on their gait; however, the algorithm can only be applied to the orthogonal view. Junhong *et al*. [10] extended this approach by separating the temporal sequence based on overlapping and non-overlapping legs by analysing the gait key frames. While ellipse fitting is improved compared to [13], the algorithm is still view dependent and requires the orthogonal view.

We expand upon the work of Lee and Grimson [13] by enabling ellipsoidal approaches to operate in the 3D domain to directly overcome limitations of view dependency and self-occlusion. We achieve this by fitting ellipsoids to a 3D voxel model and extracting features from this model to perform gait recognition. This new approach is outlined in Section 3.



(a) (b)

Figure 3. Reconstructed 3D voxel model from multi-view images.



(a) Grid based segmentation algorithm. (b) Proposed segmentation algorithm.

Figure 4. Segmentation of voxel model.

3. Volume Reconstruction and Segmentation

The algorithm presented here requires the use of multiple calibrated cameras from widely different viewpoints. Silhouettes are extracted from the video through motion segmentation (in our approach we use [5]). Projecting these silhouettes, a visual hull is formed from which a rough voxel model of the subject can be constructed. For the purpose of this paper, the z-axis is defined to be in the vertical direction, while the x-axis is defined to be in the direction of motion.

To construct the 3D human silhouette, a 3D binary volume that encapsulates the entire walking environment is built. Each point in the volume is projected in to each view to check the corresponding 2D binary silhouette at the projected point. The availability of projected points defines the 3D coordinate in the voxel volume. An example of a 3D human silhouette, created in this manner is shown in Figure 3.

For the proposed gait recognition algorithm, only features from the lower body will be considered. Similar to the approach adopted by Lee and Grimson [13], the centroid is used to separate the upper and lower body as it provides a sufficient approximation of the hip/waist. To differentiate between the thigh and lower leg, the knee is approximated to be half way between the centroid and the bottom of the model.

In the 2D case presented in [13], the left and right leg cannot be differentiated when they occlude each other. With a 3D voxel volume presented here, this is now possible. However, within a gait cycle, the legs are not constrained to move along a single plane and can pass in front (and behind) each other (as seen in Figure 3(a)). As a result, the left and right limbs cannot be separated simply along the xz plane.

In the proposed algorithm, the left/right differentiation of the upper and lower leg is performed separately. The algorithm described in Section 4.1 is used to find the major axis of the upper and lower volume distribution. We choose our segmentation plane to be orthogonal to the projection of this major axis onto the xy plane (Figure 4).

4. Gait Feature Extraction and Gait cycle segmentation

4.1. Ellipsoidal Parameter Model

Following the segmentation (see Section 3), we have four segmented regions comprising of the upper and lower legs. Extracting ellipsoidal parameters for each region involves computing the mean and covariance of the voxel distribution in that region. If x, y and z denotes the coordinates of the voxels, and the mean of the region is $(\overline{x}, \overline{y}, \overline{z})$, then the covariance matrix, Σ, is given by,

$$\Sigma = \frac{1}{N} \cdot \sum_{x,y,z} I(x,y,z) \times \begin{bmatrix} (x-\overline{x})^2 & (x-\overline{x})(y-\overline{y}) & (x-\overline{x})(z-\overline{z}) \\ (y-\overline{y})(x-\overline{x}) & (y-\overline{y})^2 & (y-\overline{y})(z-\overline{z}) \\ (z-\overline{z})(x-\overline{x}) & (z-\overline{z})(y-\overline{y}) & (z-\overline{z})^2 \end{bmatrix}, \tag{1}$$

where I is the voxel model and N is the volume of the region bounding box.

The axis orientation and lengths of the fitted ellipsoid are determined from the eigenvalues (d_1, d_2, d_3) and the eigenvectors (v_1, v_2, v_3) of the covariance matrix,

$$\Sigma \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} \begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix}. \tag{2}$$

The eigenvalues correspond to the length of each of the 3 axes, while the eigenvector is its directional vector.

The orientation of the major axis and the ratios of the axis lengths are used as the features in the classification. The axis ratios are determined by,

$$r_1 = d_2/d_1, \qquad r_2 = d_3/d_1, \tag{3}$$

while the angles used are calculated using,

$$\alpha_x = \arctan \frac{v_1 [\,0\ 1\ 0\,]}{|v_1|},$$
$$\alpha_y = \arctan \frac{v_1 [\,0\ 0\ 1\,]}{|v_1|},$$
$$\alpha_z = \arctan \frac{v_1 [\,1\ 0\ 0\,]}{|v_1|}, \tag{4}$$

assuming d_1 and v_1 are the eigenvalue and eigenvector corresponding to the major axis. From this we have five features for each region, for a total of 20 extracted features for each frame.

4.2. Gait Cycle Segmentation

Since the left and right legs are segmented separately in our proposed approach, it is easy to distinguish the left and right strides and separate each gait cycle. We define the beginning of a gait cycle as when the legs are at greatest separation and the right leg is in front. The stride length is used to segment the gait cycles as it forms a reasonably clean, cyclic signal, with peaks corresponding to the extreme strike points of each leg. This signal is smoothed using a median filter and peaks are identified using the nearest neighbour extreme minimum computation algorithm used by Sarkar *et al.* [15].

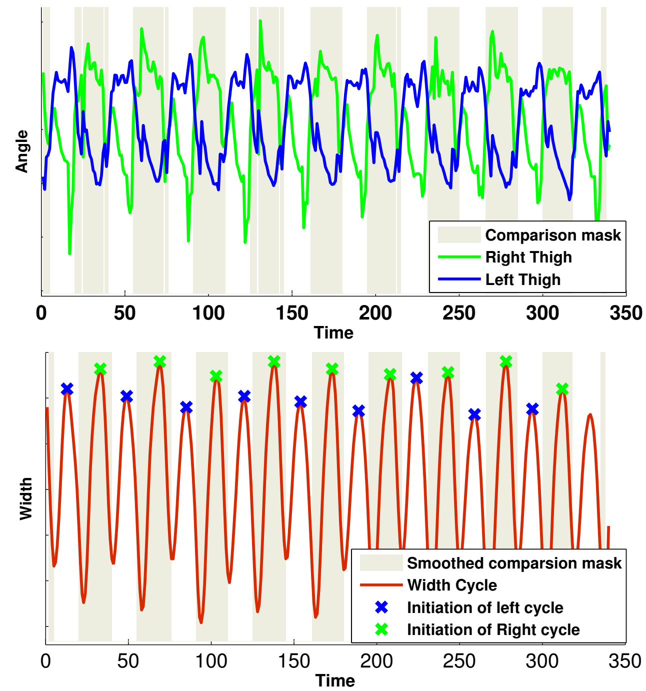


Figure 5. Gait cycle segmentation.

The y-axis angles of the upper legs are used to determine the location of the left and right legs. A logical mask is created by comparing these angles, with a smoothing window applied to remove any noise. The final mask is then used to separate the peaks as belonging to the left or right legs (Figure 5).

5. Gait Recognition and Classification

The extracted ellipsoid parameters vary throughout the gait cycle and form a signal. To obtain our final feature vector, we extract harmonic components from these signals. For each ellipsoid parameter, we apply a Discrete Fourier Transform over a single gait cycle. We take the first 3 components, leaving them as complex pairs, as the final features used in the classification. With 20 parameters, the resultant feature vector for each gait cycle is of length 60.

Before extracting the Fourier harmonics, the features are scaled such that they range between 0 and 1 in the gallery set. This scaling is then similarly applied to incoming probe sequences.

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

$$d_{ij} = \sqrt{\sum |F_i - F_j|^2}, \tag{5}$$

remembering that the feature values are complex.

To determine the distance between two sequences, each of which is composed of multiple gait cycles, we use the

Experiment	Gallery Type	Probe Type
Exp.2a	Slow Walk (SW)	Ball (B)
Exp.2b	Slow Walk (SW)	Fast Walk (FW)
Exp.2c	Ball (B)	Fast Walk (FW)

Table 1. Experiment 2 (Inter-class test cases).

algorithm proposed by Boulgouris *et al*. [2]. From this, the distance to the closest gallery cycle from each probe cycle ${d_{\min}}_i^P$, and the distance to the closest probe cycle from each gallery cycle ${d_{\min}}_j^G$ is found,

$$ {d_{\min}}_i^P = \min_j \left(d_{ij}\right), \qquad {d_{\min}}_j^G \left(d_{ij}\right). \tag{6} $$

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

$$ D = \frac{1}{2}\left(\text{median}\left({d_{\min}}^P\right) + \text{median}\left({d_{\min}}^G\right)\right). \tag{7} $$

6. Experiments and Results

The performance of our algorithm is evaluated using the CMU MoBo database [7]. It includes 25 subjects captured from 6 camera views with more than 8 gait cycles per subject in four types of walking conditions; slow walk, fast walk, walking while carrying a ball, and walking on an inclined surface. We test both intra-class classification using slow walk, fast walk, and walking while carrying a ball (labelled Exp.1a-1c); and inter-class classification as outlined in Table 1 (labelled Exp.2a-2c).

In intra-class tests, the number of gait cycles for each subject is split in two, the first half is used for the gallery and the second half for the probe. Similarity distance values are computed as explained in Section 5 between each probe and gallery subject and performances is evaluated using ROC (Receiver Operating Characteristics) curves. For our baseline, we use an algorithm similar to Lee and Grimson [13] applied to one of the side views. Ellipses are fitted to the silhouette for the seven segmented regions as in [13], however only the angle and axis ratio of each ellipse are used. Classification is identical to that of Section 5 with harmonics extracted from each of the 14 parameters. Results are shown in Figure 6.

As expected, intra-class tests resulted in high classification rates for both approaches. However, the proposed algorithm outperformed the baseline, achieving a 100% verification rate at a false alarm rate of 10% for all cases (see Table 2 and Figure 6(a) - 6(b).

For the inter-class test cases, the proposed algorithm significantly outperforms the baseline. From Figure 6(d) - 6(f), it can be seen that the proposed algorithm outperforms the baseline at all operating points. At an operating point of 10% FAR, between 15-20% improvement in verification rate is achieved. This improvement can be directly attributed to the algorithm's ability to bypass the problem of self-occlusion, which hampers the original 2D method's ability to accurately model the gait dynamics.

Experiment	Proposed Method	Baseline Method
Exp.1a (SW)	100	93.8
Exp.1a (B)	100	91.0
Exp.1a (FW)	100	99.5
Exp.2a	70.5	50.5
Exp.2b	78.6	63.3
Exp.2c	61.0	42.4

Table 2. Verification rate at FAR of 10%.

7. Conclusions and Future Work

In this paper, we have presented a novel gait recognition algorithm using a 3D voxel model, derived from silhouettes from multiple views, to extract gait features based on ellipsoid parameterisation of the voxel model. The use of 3D information allows left and right legs to be easily segmented, and gate cycles to be detected. The proposed approach achieves a significant performance increase over its 2D counterpart, particularly when there is a class mismatch between the gallery and probe sequences.

Future work will focus on improving the segmentation to ensure that parts of the upper body, such as hands, do not interfere with the extracted regions, and extracting additional information such as the feet. Further evaluations will also be carried out on more challenging data (such as surveillance footage) to further illustrate the validity of the proposed approach.

References

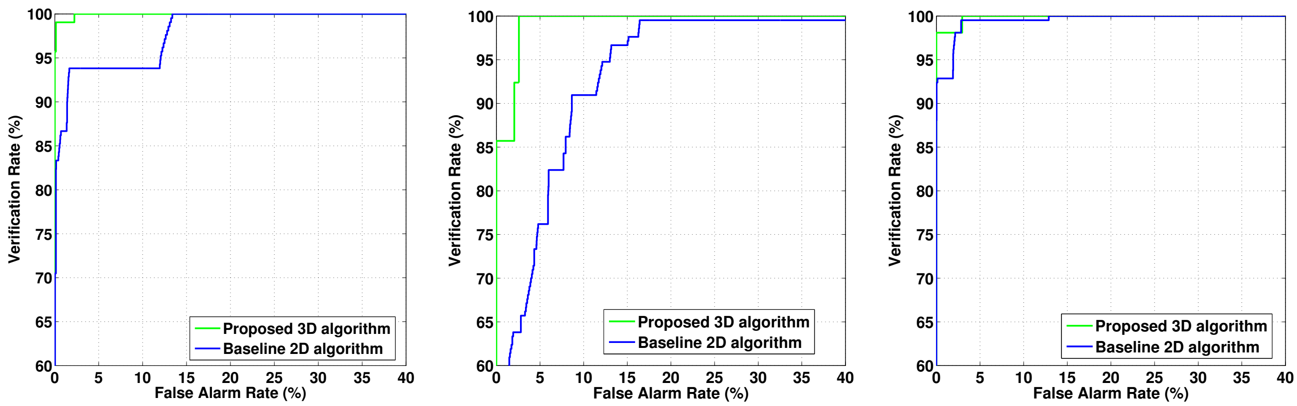
[1] C. BenAbdelkader, R. Cutler, and L. Davis. Motion-based recognition of people in eigengait space. In *Proc. IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pages 267–272, 2002.

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

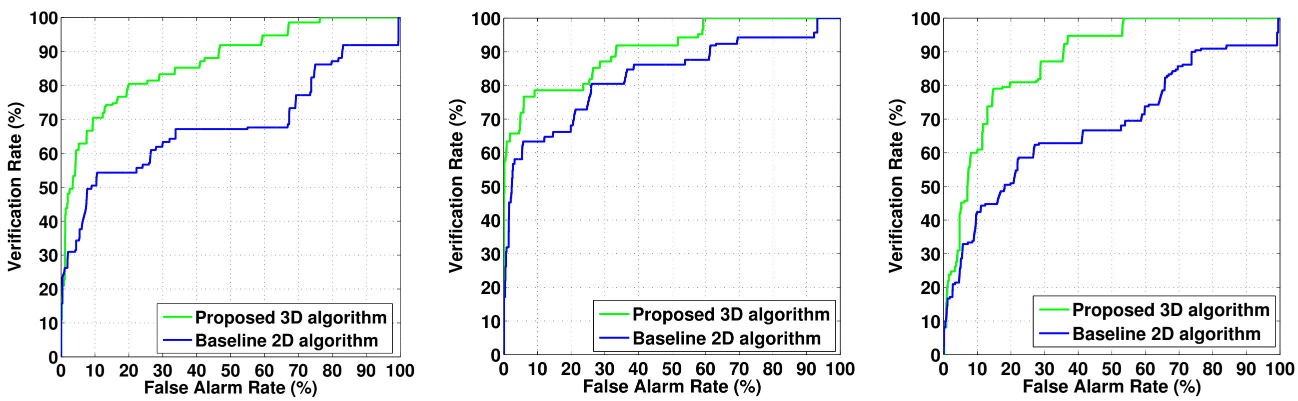
[3] J. E. Boyd and J. J. Little. Biometric gait recognition. *Advanced Studies in Biometrics*, pages 19–42, 2005.

[4] 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.

[5] S. Denman, C. Fookes, and S. Sridharan. Improved simultaneous computation of motion detection and optical flow for object tracking. In *Proc. Digital Image Computing: Techniques and Applications*, pages 175–182, 2009.



(a) Exp.1a : Slow Walk (b) Exp.1b: Ball (c) Exp.1c: FastWalk



(d) Exp.2a (Gallery: SW, Probe: B) (e) Exp.2b (Gallery: SW, Probe: FW) (f) Exp.2c (Gallery: B, Probe: FW)

Figure 6. Recognition results.

[6] C. Fookes, S. Denman, R. Lakemond, D. Ryan, S. Sridharan, and M. Piccardi. Semi-supervised intelligent surveillance system for secure environments. In *Proc. IEEE Int. Sym. on Industrial Electronics*, pages 2815 –2820, 2010.

[7] 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.

[8] Z. Guoying, L. Guoyi, L. Hua, and M. Pietikainen. 3d gait recognition using multiple cameras. In *Proc. IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pages 529–534, 2006.

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

[10] X. Junhong, C. Wang, L. Jin, W. Lei, L. Lijie, and L. Hong. Gait recognition based on key frame and elliptical model. In *Proc. IEEE Int. Conf. on Information and Automation*, pages 2483–2487, 2010.

[11] G. Junxia, D. Xiaoqing, W. Shengjin, and W. Youshou. Action and gait recognition from recovered 3-d human joints. *IEEE Trans. on Systems, Man, and Cybernetics*, 40(4):1021–1033, 2010.

[12] 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 Trans. on Image Processing*, 13(9):1163–1173, 2004.

[13] L. Lee and W. E. L. Grimson. Gait analysis for recognition and classification. In *Proc. IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pages 148–155, 2002.

[14] M. S. Nixon and J. N. Carter. Advances in automatic gait recognition. In *Proc. IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pages 139–144, 2004.

[15] 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.

[16] D. K. Wagg and M. S. Nixon. Model-based gait enrolment in real-world imagery. In *Proc. Multimodal User Authentication*, pages 189–195. University of California, Santa Barbara, 2003.

[17] C. Xiang-tao, F. Zhi-hui, W. Hui, and L. Zhe-qing. Automatic gait recognition using kernel principal component analysis. In *Proc. Int. Conf. on ICBECS*, pages 1–4, 2010.

[18] E. Zhang, Y. Zhao, and W. Xiong. Active energy image plus 2dlpp for gait recognition. *Signal Processing*, 90(7):2295 – 2302, 2010.