

# Robust Arbitrary-View Gait Recognition Based on 3D Partial Similarity Matching

Jin Tang, Jian Luo, Tardi Tjahjadi, *Senior Member, IEEE*, and Fan Guo

**Abstract**—Existing view-invariant gait recognition methods encounter difficulties due to limited number of available gait views and varying conditions during training. This paper proposes gait partial similarity matching that assumes a 3D object shares common view surfaces in significantly different views. Detecting such surfaces aids the extraction of gait features from multiple views; 3D parametric body models are morphed by pose and shape deformation from a template model using 2D gait silhouette as observation. The gait pose is estimated by a level set energy cost function from silhouettes including incomplete ones. Body shape deformation is achieved via Laplacian deformation energy function associated with inpainting gait silhouettes. Partial gait silhouettes in different views are extracted by gait partial region of interest elements selection and re-projected onto 2D space to construct partial gait energy images. A synthetic database with destination views and multi-linear subspace classifier fused with majority voting is used to achieve arbitrary view gait recognition that is robust to varying conditions. Experimental results on CMU, CASIA B, TUM-IITKGP, AVAMVG, and KY4D data sets show the efficacy of the propose method.

**Index Terms**—Gait, person identification, 3D gait model, partial similarity matching.

## I. INTRODUCTION

GAIT is difficult to be disguised and can be easily observed in low-resolution video sequences at long distance without contact with the human subject. Thus gait recognition has great potential, and could be useful in various scenarios, such as analysis of forensic evidence for a legal case [1], criminal investigation and security check [2]. It could be achieved with high recognition rate where environmental and certain factors are controlled to some extent. The gait database OU-ISIR LP with 4000 subjects [3] has been used to evaluate the upper bound accuracy of gait recognition. However, implementing a gait-based personal identification

Manuscript received March 31, 2016; revised July 23, 2016; accepted September 6, 2016. Date of publication September 22, 2016; date of current version November 15, 2016. This work was supported in part by the National Natural Science Foundation of China under Grant 61403426 and Grant 91220301, in part by the Natural Science Foundation of Hunan, China, under Grant 2016JJ2150, and in part by the Science and Technology Key Program of Hunan, China, under Grant 2015WK3006. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Lei Zhang. (*Corresponding author: Jian Luo*.)

J. Tang and F. Guo are with the School of Information Science and Engineering, Central South University, Changsha 410083, China (e-mail: tjin@csu.edu.cn; guofancsu@163.com).

J. Luo is with the School of Information Science and Engineering, Central South University, Changsha 410083, China, and also with the College of Physics and information Science, Hunan Normal University, Changsha 410000, China (e-mail: delphifx@csu.edu.cn).

T. Tjahjadi is with the School of Engineering, The University of Warwick, Coventry, CV4 7AL, U.K. (e-mail: t.tjahjadi@warwick.ac.uk).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIP.2016.2612823

system for large population in a real application is still challenging due to the imaging of natural gait being easily influenced by various factors, such as scene structure, illumination, clothing variation, footwear, carrying conditions, walking speed, occlusions and camera views [4]–[6]. Nevertheless, gait provides an additional biometric cue for reliable personal identification system.

In order to obtain acceptable recognition rate, it is important to obtain good segmentation of the gait silhouettes. However, many factors often lead to inaccuracy in the segmentation, e.g., similarity of colours between the subject's clothes and background, changes in background due to illumination or appearance of new objects, self occlusion, etc. [7]. Gait image inpainting methods have thus been proposed, i.e., partial derivative equations based, texture synthesis, semi-automatic inpainting and hybrid methods. Multiple filters have been introduced to reduce noise or enhance the image quality. Image matching and moving object detection have been used to address dark areas of the data and to perform hole-filling. However, these methods assume that the missing pixels are similar to their adjacent pixels and thus use neighbouring pixels for interpolation. When the missing data is large, e.g., due to significant occlusion, existing model-free inpainting methods fail to correctly segment the gait silhouettes. Other methods for addressing imperfect gait silhouettes use robust statistics to extract gait features in a static image by averaging the silhouettes in a gait period. The gait energy image (GEI) [8] is one such method that is less sensitive to uncorrelated silhouette noise. Other variants include Gait Flow Image (GFI) [9], Pose Kinematics and Pose Energy Image (PEI) [10] and Frame Difference Energy Image (FDEI) [7]. However, these energy images are mainly for mitigating the effect of imperfect silhouettes in covariate conditions that affect gait feature extraction rather than the gait itself. Thus, in this paper we introduce a gait inpainting method that uses the unified and prior knowledge of 3D human body to restore the incomplete gait silhouettes when there is large missing data.

Existing 2-dimensional (2D) gait recognition systems encounter difficulties in multi-view gait recognition where there is a limited number or absence of some gait views in the training data. Since gait is 3-dimensional (3D), a 2D gait image sequence provides only single-view information that constrains gait recognition from an arbitrary view. Using 2.5 dimensional (2.5D) or 3D models for gait recognition [11], [12] has achieved significant advances due to the availability of 3D imaging devices where the data is used for 3D reconstruction and target tracking. However the use of inexpensive 3D

sensors has its limitations, e.g., low-resolution and unreliable coarse range data for dark and infrared-absorbing objects, missing data if the objects are out of the sensor's range, and data flickering in successive frames. Thus, it is difficult to use only range data to accurately segment body contours. The coarse data of the head and feet are also sometimes missing. As a result the 2.5D gait voxel models derived from the coarse range images are also imperfect. The use of such an approach provides only partial view of the gait data, similar to the use of a 2D camera. It only provides a simplified 3D surface representation that contains one range value ( $z$ ) for every point in the  $(x,y)$  image plane. The use of high-accuracy 3D whole body scanners are typically expensive and limited for static modelling due to the long scanning time, e.g., for estimation of the body shape beneath the clothes.

For view-invariant gait recognition to be practical, few cameras should be used. In this paper, we use 2D gait images that include incomplete body images to estimate the 3D gait model. The 3D gait pose is first estimated using level set energy cost function and improved using Laplacian deformation. The method not only repairs the incomplete 2D gait images but also estimates its corresponding 3D model. Since only one-view gait images are used, the 3D model is imperfect. Thus, a gait partial similarity matching (GPSM) is proposed for extracting the partially matched gait features in a cycle from 3D gait models with different views. These are the common data of a surface viewed from significantly different camera locations, referred to as the common view surface. Detecting the common view surfaces in different views and transforming the surfaces to the same view aids the extraction of the gait features from multiple views. We refer this as gait partial similarity matching and use it to achieve gait recognition from different views, especially for probe data from view that is not in the training dataset, i.e., gait recognition from arbitrary views but with limited number of gait views available during training.

The contributions of this paper are: (1) use of 3D human pose estimation and shape deformation to reconstruct parametric 3D body from 2D data to achieve robustness to variation in significant view changes, and occlusions, e.g., missing body parts or segmentation noise; (2) use of 2D incomplete gait image inpainting with weighted level set function based on 3D body to significantly improve the quality of incomplete 2D gait silhouettes for improved recognition; (3) by using partial similarity matching, view-invariant 2D gait images can be matched directly by partial similar features enabling robustness to carried item and improved recognition rate; and (4) by introducing multi-linear subspace classifier with majority voting for fusion enables arbitrary views gait recognition using GPSM.

This paper is organized as follows. Section II presents the related work. Section III discusses the gait partial similarity matching. Section IV presents our multi-view gait recognition method based on multi-linear subspace classifier with majority voting. Section V presents the experimental results and Section VI concludes the paper.

## II. RELATED WORK

There are two main approaches to view-invariant gait recognition: (1) model-based human body structural analysis and

(2) model-free appearance-based statistical analysis. The first approach uses different 3D gait skeleton structural models for view-invariant feature extraction of length of torso, upper arm, lower arm, thigh, calf and foot [13]. Zhao et al. [14] explored the use of multiple cameras to construct skeleton model with 10 joints and 24 degree of freedom. The lengths of key segments are extracted as static parameters, and their motion trajectories used as dynamic features to realize 3D gait recognition. Rogez et al. [15] modelled 3D body poses and camera viewpoints with a low dimensional manifold, and use it to learn a generative model of the silhouette for view invariant 3D gait tracking. However all current modelling of 3D gait skeleton structural models is mostly based on images from multiple cameras, and are thus inaccurate. Also the features extracted from these models are limited.

The second approach uses 3D appearance, 2D image appearance or 2.5D depth image appearance. Shakhnarovich et al. [16] used an image-based visual hull to construct a 3D model that is rotated to realize view-invariant gait recognition by projecting 3D surface to 2D space. Sivapalan et al. [17] used a 3D volume model derived from multi-view 2D images or frontal depth images, and the back-filled versions for matching frontal-view depth image and side-view 2D silhouette for cross-modal cross-view gait recognition. However, the need of complex hardware setup makes it practical only under limiting laboratory setting.

There are three approaches to 2D appearance-based gait recognition. The first adopts view normalization or uses the result fused from multiple view-dependent gait recognition. Nizami et al. [18] used Extreme Learning Machine multiclass classifier where the view-dependent classifications are fused at score level according to some rules. This method is limited by the size of the multi-view gait databases and cannot realize arbitrary-view gait recognition. Jean et al. [19] used view-normalized body part trajectories as view-invariant gait features, but these are impractical for significantly different views and when self-occlusion is encountered. Wei and Cong [20] introduced the deterministic learning theory to achieve view-invariant gait recognition and extraction of gait silhouette features that represents gait dynamics and reflects view variation.

The second uses projection relationship of gaits across different views. Makihara et al. [21] proposed a view transformation model (VTM) to transform a probe gait view data onto the virtue view that exists in the training database. Muramatsu et al. [22] proposed arbitrary gait view transformation scheme using 3D gait data via VTM. Kusakunniran et al. [23], [24] developed a VTM by using correlated motion regression and multi-layer perceptron. However a VTM based method requires multi-view 2D images to construct the VTM and performs virtue view transformation during recognition. Also, the parameters of the VTM are sensitive to the training multi-view images.

The third relies on correlation learning mapping relationship of gaits across multi-views. Hu et al. [25] proposed view-invariant discriminative projection to improve the discriminative ability of multi-view gait features using a unitary linear projection. The multi-view gait features are matched without knowing or estimating the viewing angles. Hu [26] used a

sparse local discriminant canonical correlation analysis (CCA) to model the correlation of gait features from different views, using the correlation strength as similarity measure. Multi-view gait features correlation analysis does not require the gait to be transformed to virtue view. However, the gait features are less correlated when the gait views are significantly changed, and multi-view gait images are needed in the training step.

Numerous gait recognition methods have been proposed to address multiple covariates (e.g., occlusion, clothing invariant, carrying items, and unconstrained paths). Martín-Félez and Tao [5] treat gait recognition as a bipartite ranking problem. By exploiting learning to rank, the method is suitable under an uncooperative setting and robust against changes in covariate conditions. To identify subjects walking along curved trajectories Iwashita *et al.* [27] used a 4D gait database to synthesize virtual image corresponding to the estimated direction, and affine moment invariants as gait features. Rida *et al.* [28] select the most discriminative body part based on group Lasso of motion to reduce the intra-class variation. Xin *et al.* [29] used integrated sparse coding and multi-view hypergraph learning for recognizing a pedestrian under uncooperative setting. López-Fernández *et al.* [30] presented a rotation invariant gait descriptor based on 3D angular analysis of the subject's movement for multi-view gait recognition on unconstrained paths. Nandy *et al.* [31] used pooled segmented statistical features to describe the shape of GEI edge contour. The higher order moments of the shape centroid in conjunction with a set of other area based features are combined to improve the classification results. Most of these methods address different covariates with limited gait views, mainly using frontal or side view of the gait sequences. In fact, it is difficult for a 2D training dataset to cover all conditions, especially incomplete gait silhouettes and with different carrying that affect the overall body shape directly.

To address the above-mentioned gait recognition problems with arbitrary views, we propose partial similarity matching based on reconstructed 3D gait model with inpainting. A single view gait image is used to estimate the 3D gait model that includes one-side virtual surface portion of the human body in a gait cycle. The partial gait silhouettes in different views are extracted using gait partial region of interest (ROI) elements selection method and re-projected onto 2D space to construct partial gait energy images (PGEIs). A synthetic database with destination views is constructed based on GPSM, and multi-linear subspace classifier fused with major voting is used for gait classification and subject identification.

### III. GAIT PARTIAL SIMILARITY MATCHING

#### A. Overview

GPSM in 3D space enables gait recognition from different views especially from arbitrary views using limited number of gait views available in training data as illustrated in Fig.1. In Euclidean space, given two different 3D surface models  $S_1$  and  $S_2$ , the 3D partial surface matching searches for the optimal transformation  $T = \{R, t\}$ , where  $R$  is rotation matrix

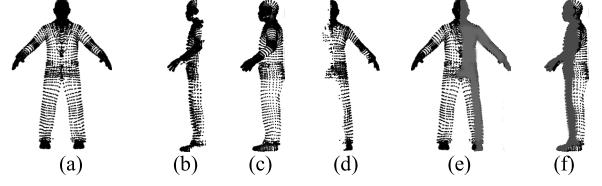


Fig. 1. Gait partial similarity matching: (a) mesh vertices at  $0^\circ$  view (i.e., front view); (b)  $90^\circ$  view partial mesh vertices obtained by removing the self-occlusion data in (a) and rotated to  $90^\circ$ ; (c) mesh vertices at  $90^\circ$  view; (d)  $0^\circ$  view partial mesh vertices obtained by removing the self-occlusion data in (c) and rotated to  $0^\circ$ ; (e) grey part of body denotes the partial similarity matching of (a) and (c) at  $0^\circ$  view using (d) as partial feature; and (f) grey part of body denotes the partial similarity matching of (a) and (c) at  $90^\circ$  view using (b) as partial feature.

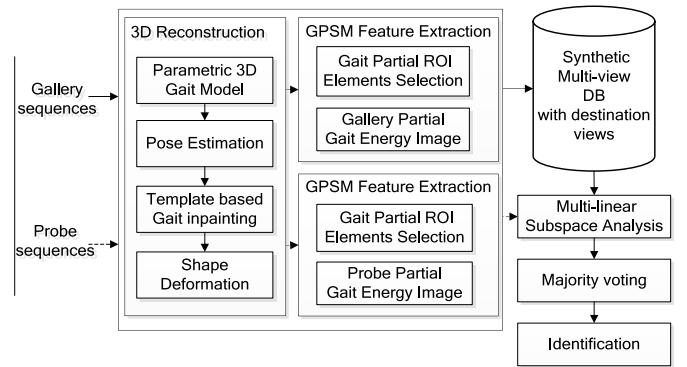


Fig. 2. Overview of partial similarity matching for gait recognition.

and  $t$  is translation matrix, by solving

$$\arg \min_T \sum_{(v, v') \in C} \|Rv + t - v'\|^2 \quad (1)$$

where  $C$  denotes the dataset of GPSM points between the two surfaces. Most methods for solving 3D partial similarity matching, e.g., Besl [32] and Bareque [33], use two static 3D images of the same object captured at different views, and a rigid transformation  $T$ . As a gait cycle is composed of several dynamic non-rigid body transformation models, it is difficult to just use  $T$  to match the training subjects with themselves or with each other. Thus, in this paper a partial similarity matching (based on selected gait partial ROI elements between the given views) and PGEIs as features are used to realize arbitrary view gait recognition.

The overview of the partial similarity matching is shown in Fig. 2. In order to reconstruct the 3D gait model with different poses, an articulated skeleton structure is embedded into the body mesh, and a level set energy cost function is used for estimating the gait pose from incomplete 2D gait silhouettes. Laplacian deformation is then performed and the estimated pose mesh fitted onto the detailed body shape using the repaired 2D gait silhouette contour as reference. The gait frames used for 3D reconstruction are captured by a single view camera and include only the front view of the human body as silhouette constraints. Due to the absence of the side or rear surfaces, GPSM is used for partial similar feature extraction and representation. The multi-view PGEIs are then obtained to form a novel synthetic gallery database for

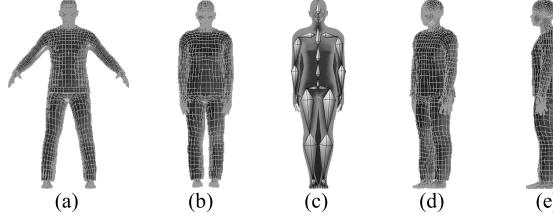


Fig. 3. 3D body template mesh: (a) T-pose 3D parametric body mesh; (b) initialized I-pose body mesh at 0° view; (c) skeleton embedded in I-pose model mesh; (d) I-pose model at 45° view; and (e) I-pose model at 90° view.

multi-view gait training. Multi-linear subspace analysis with majority voting is used for subject identification.

### B. 3D Gait Model Reconstruction

In 2D space, due to the absence of range value for every point in the image plane, partial similarity matching features cannot be extracted directly. Also, most research on 3D human motion capture assumes the body shape is known and is represented coarsely (e.g., using cylinders or super quadrics to fit limbs, or using visual hull without skeleton embedded). Such representations are less useful for 3D gait recognition with covariates. To address these problems, a parametric 3D human body is chosen as a template model obtained using the software Makehuman [34] that incorporates 1170 morphings. The parametric body model is optimized for subdivision surfaces modelling with 15128 vertices, suitable for gait modelling. Since gait silhouettes are often imperfect, by using parametric template, the model based gait silhouette inpainting can be derived for 3D gait model reconstruction.

The reconstruction of 3D gait model from incomplete 2D gait silhouettes involves the following steps: 1) initialize the 3D parametric body template; 2) estimate 3D body pose; 3) template based 2D incomplete gait image inpainting; and 4) 3D shape deformation using the complete 2D gait silhouette.

*1) Initializing 3D Parametric Body Template:* A standard T-pose 3D parametric body is derived from the Makehuman software with average semantic values as shown in Fig.1 (a). Since T-pose body is not well fitted with the human postures in a gait cycle, a model with an I-pose as shown in Fig.1 (b) is initialized as 3D body template by morphing T-pose model onto I-pose using skeleton-based mesh deformation method [35]. The template model mesh is parameterized by the vector  $X = \{V_X, T_X\}$  comprising  $M$  vertices  $V_X = \{v_1, v_2, \dots, v_m, \dots, v_M\}$  and  $I$  mesh faces  $T = \{t_1, t_2, \dots, t_I\}$ . We use an articulated skeleton structure from [36] as shown in Fig. 3(c).

It is assumed that the surface of template  $X$  comprises the set of  $P = \{p_1, p_2, \dots, p_j, \dots, p_J\}$  rigid parts that are associated with  $J$  joints of the articulated skeleton. Every vertex in the template model is associated with a part label  $\eta_j$ . It denotes the rigid part to which the vertex belongs. Every rigid part  $p$  is associated with a set of vertices and it has the same set of transformations in pose deformation, i.e., the joint rotation. The rigid parts of the human body are associated with the body joints and are estimated using the algorithm in [37].

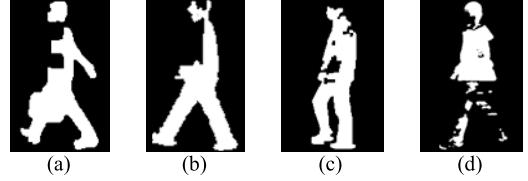


Fig. 4. 2D incomplete gait silhouettes from CASIA database.

*2) Estimation of 3D Body Pose Based on Energy Cost Function:* Most recent work on body pose estimation and tracking using image cost function or other Bayesian methods require a generative model. However, crude and structured models are often used, e.g., articulated body model or ellipsoidal parametric model [38]. The method for recovering 3D pose throughout an image sequence by using SCAPE parametric body models in [39] based on 2D images assumes that the level of detailed shape recovery can be improved with additional cameras and improved background subtraction.

Let  $S(x, y)$  be the RGB frame captured by a fixed camera. Binary human silhouette image  $\bar{S}(x, y)$  is extracted using the background subtraction algorithm in [40] which can cope with local illumination changes, such as shadows and highlights, as well as global illumination changes. However, gait silhouettes extracted from a complex environment might still be incomplete, e.g., those in Fig. 4, where large areas of the silhouette are missing, mainly due to the similarity of colours between the subject's clothes and background. It is thus important to estimate the detailed 3D body shape and pose directly from the incomplete images.

In this paper, the observation 2D gait silhouette images (complete or incomplete) are used to estimate 3D body poses. Let the estimated 3D gait model be denoted by  $Y = \{V_Y, T_Y\}$  and  $V_Y = \{y_1, y_2, \dots, y_n, \dots, y_M\}$  where  $M$  is the number of vertices. The instance model  $Y$  is obtained from the template model  $X$  with pose deformation. The body pose is associated with the joint angles of model skeleton. The vertices of the new pose deformed model  $Y$  is denoted by

$$V_Y = \{v_m \cdot R(\Delta\alpha_{\zeta(m)})\}, \quad (2)$$

where  $\Delta\alpha$  is joint relative rotation angles,  $m = 1 \dots M$ ,  $v_m$  are vertices in  $V_X$  of template model,  $R$  is rigid part transform matrix,  $\zeta(m) \in \eta_j$ ,  $j = 1, 2, \dots, J$  determines to which rigid part  $v_m$  belongs,  $\eta_j$  is part label, and  $\psi = [\Delta\alpha_{\eta_1} \Delta\alpha_{\eta_2} \dots \Delta\alpha_{\eta_J}]$  which is denoted by the joints rotation matrix from the template I-pose model. Using the joint relative rotation angles, a hierarchical skeleton with rigid bones connected by joints is constructed. The 3D mesh model is skinned to the given posture using the method in [35], i.e., the template 3D I-model is morphed onto any given posture.

Level set was introduced in [41] for capturing moving fronts to address image segmentation problems. We do not use the level set directly for gait contour segmentation from complicated background images. Instead, we apply level set algorithm to incomplete 2D gait silhouettes to construct pose energy cost function,  $E_t$ , image inpainting and gait image inpainting with 3D templates for estimation of body pose.

Let the moving active contours or moving front be denoted by the zero level set  $\zeta(t) = \{(x, y), \phi(t, x, y) = 0\}$  where

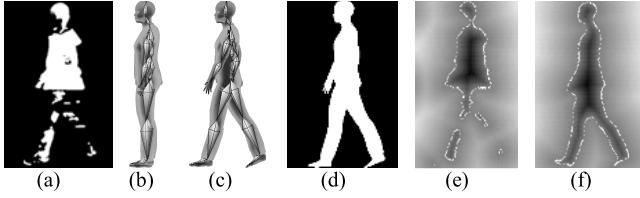


Fig. 5. Level set of gait images: (a) 2D incomplete gait image; (b) I-pose 3D body template mesh with skeleton embedded; (c) estimated 3D gait model by posture morphing from (b); (d) 2D projected gait image from morphed mesh (c); (e) level set of (a) in X-Y plane; and (f) level set of (d) in X-Y plane.

$\phi(t, x, y)$  is the level set function at time  $t$ . The evolution equation of level set function is

$$\frac{\partial \phi}{\partial t} + F |\nabla \phi| = 0, \quad (3)$$

where  $F$  denotes the speed function. In image segmentation  $F$  is associated with the level set function  $\phi$  and the gradient of the image data. The level set function at time  $t$  is usually defined as a signed distance function

$$\phi(\vec{r}(x, y), t) = \begin{cases} d(\vec{r}, \zeta(t)) & \vec{r} \in \Omega^+ \\ 0 & \vec{r} \in \zeta(t) \\ -d(\vec{r}, \zeta(t)) & \vec{r} \in \Omega^- \end{cases} \quad (4)$$

where  $\vec{r}(x, y)$  is the spatial vector determined by the point  $(x, y)$  in 2D image plane, and  $d(\vec{r}, \zeta(t))$  denotes the signed distance between vector  $\vec{r}$  and the zero level set active contour  $\zeta(t)$  at time  $t$ . The distance function  $d(\vec{r}, \zeta(t))$  is defined as  $d(\vec{r}, \zeta(t)) = \min(|\vec{r} - \vec{r}_I|)$ , where  $\vec{r}_I \in \zeta(t)$ .  $\Omega^+$  denotes the area outside the active contour while  $\Omega^-$  is inside.

The level set function is re-initialized periodically to be a signed distance function during evolution by solving

$$\frac{\partial \phi}{\partial t} = \text{sign}(\phi_0)(1 - |\nabla \phi|). \quad (5)$$

Since the re-initialization is quite complex, we use the variational level set formulation of curve evolution without re-initialization [42], i.e.,

$$\frac{\partial \phi}{\partial t} = \mu \left[ \Delta \phi - \text{div}\left(\frac{\nabla \phi}{|\nabla \phi|}\right) + \lambda \delta(\phi) \text{div}\left(g \frac{\nabla \phi}{|\nabla \phi|}\right) + \nu g \delta(\phi) \right] \quad (6)$$

where  $\lambda > 0, \mu$  and  $\nu$  are constants. The edge indicator function

$$g = \frac{1}{1 + |\nabla G_\sigma * I|^2}, \quad (7)$$

where  $G$  is the Gaussian kernel with standard deviation  $\sigma$ . The approximation of (6) is expressed as [42]

$$\phi_{i,j}^{k+1} = \phi_{i,j}^k + \tau L(\phi_{i,j}^k), \quad (8)$$

where  $L(\phi_{i,j}^k)$  is an approximation of the right hand side of (6) by the above spatial difference scheme.

The performance of level set depends on how its parameters are set, especially for complex background. We use the level set method to process incomplete gait silhouettes, e.g., Fig. 5(a), and construct level set images, e.g., Fig. 5(e)-(f). For our gait silhouettes, we used the parameters  $\lambda = 5.0$ ,

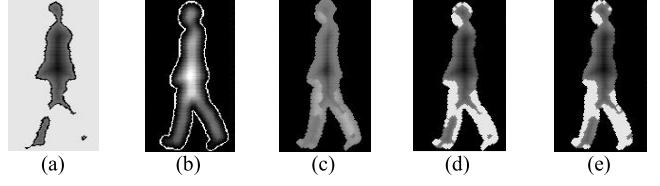


Fig. 6. Level set energy cost function construction: (a) processed level set of 2D incomplete gait image; (b) 3D projected gait silhouette after weighting; (c) mixture of the two data sets of (a) and weighted silhouette (b); and (d) and (e) are mixture of the data set of (a) and two different pose projected silhouettes without weighting.

$\mu = 0.04$ ,  $\nu = 3.0$ , and time step  $\tau = 5.0$  as similarly used in [42], which is significantly larger than the time step used for traditional level set methods. The curve evolution takes 300 iterations.

Let the observation 2D gait image shown in Fig. 5(a) be denoted by  $I_{2D}(x, y)$  and the corresponding 3D to 2D mapping gait image be  $I_{Y_t}^\theta(x, y)$  as illustrated in Fig. 5(d), which is derived from the estimated 3D gait model  $Y_t$  at time  $t$ .  $Y_t$  as illustrated in Fig. 5(c) is morphed from template I-pose gait mesh model using human skeleton skinned mesh animation method. The level set method is used to obtain the silhouettes of the two gait images, where the final evolution level sets are denoted by  $\phi_{2D}$  and  $\phi_{3D}$ . The silhouette contours are denoted by  $\zeta_{2d}^0 = \phi_{2D}(x, y) = 0$  and  $\zeta_{3d}^0 = \phi_{3D}(x, y) = 0$ . The data inside the silhouette contours have negative value  $\phi < 0$  while the data outside the contours have positive value  $\phi > 0$  as illustrated in Fig. 5.

Let the level set of the observation 2D gait image  $\phi_{2D}^1 = \phi_{2D} H(-\phi_{2D}) + H(\phi_{2D})$  be 1 where  $\phi_{2D} > 0$  and retain the value unchanged for  $\phi_{2D} \leq 0$  as shown in Fig. 6(a), where  $H(\cdot)$  is the Heaviside function. The gait data inside the level set contour of 3D projected gait image is extracted and denoted by  $S_{3D} = H(-\phi_{3D})$ . The data set inside the level set contour are then weighted by

$$D_w(x, y) = G_\sigma * d(\vec{r}, \zeta_{3d}^0), \quad (9)$$

where  $d(\vec{r}, \zeta_{3d}^0)$  denotes the distance between vector  $\vec{r}(x, y)$  and curve  $\zeta_{3d}^0$ .  $G_\sigma$  is the Gaussian kernel with standard deviation  $\sigma$ .  $*$  denotes the convolution operation and  $(\cdot)$  denotes the dot product between two matrices. Let  $\phi_{mix} = \phi_{2D}^1 \cdot (D_w \cdot S_{3D})$  which is the weighted mixture of the two level set shown in Fig. 6(c).

The mixed level set  $\phi_{mix}$  is then segmented into three parts, i.e., head, torso and motion leg. The anatomical positions of neck, hip and knee as a fraction of the body height ( $H$ ) following anatomical studies in [43] as 0.870H, 0.530H and 0.285H, respectively, measured from the bottom of the bounding rectangle are used for segmentation. Note that these positions, which are based on anthropometry, might slightly deviate from the actual positions on a subject especially when the subject is performing an activity, e.g., walking. For more accurate lower body segmented parts associated with the most important body motion, the human leg motion is modeled as a pendulum [44]. The fixed point of the pendulum in a gait cycle is selected as the segmented line. Since partitioning a

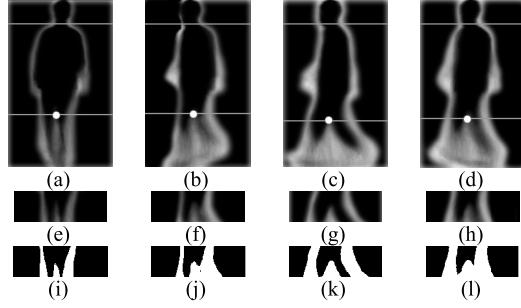


Fig. 7. Segmentation of sub-level set parts using GEnIs: (a)-(d) GEnIs in 18°, 54°, 90°, 126° views with horizontal lines as estimated positions of neck and fixed point of pendulum; (e)-(h) segmented parts between hip and knee from (a)-(d); and (i)-(l) corresponding binary images of (e)-(h).

body according to its anatomical position is still challenging, the human silhouettes are sometimes incomplete. To overcome this, the frames of gait silhouettes in a cycle are first averaged using GEnI [45] as illustrated in Fig 7(a)-(d) without the horizontal lines. The lower part of GEnI between anatomical positions hip and knee are segmented as shown in Fig. 7(e)-(h) and converted to binary images, i.e., Fig. 7(i)-(l). Using the corresponding binary images, the fixed point of pendulum is estimated from the position of the second peak.

Let  $H_s = 0.870H$  be the anatomical position of neck. The estimated position, which might be closer to the actual location, is defined as  $\hat{H}_s \in [H_s - \Delta h, H_s + \Delta h]$ , where  $\Delta h$  denotes the interval parameter. Let  $L_{\hat{H}_s}$  denotes the slice width from left to right body contour in  $\hat{H}_s$  horizontal position. The value of  $\hat{H}_s$  is estimated by  $\hat{H}_s = \arg \min_{\hat{H}_s \in [H_s - \Delta h, H_s + \Delta h]} |L_{\hat{H}_s}|$ .

Let the sub level set body parts be denoted by  $\phi_{mix}^n$  where  $n=1, 2, 3$ . The level set energy cost function is

$$E_t = - \sum_{i=1}^3 \mu_{neg}^i E_{neg}^i + \mu_{mix} E_{mix}, \quad (10)$$

where  $E_{neg}^i = \|H(-\phi_{mix}^i) \cdot \phi_{mix}^i\|_{L_1}$  and  $E_{mix} = \sum_x \sum_y (\phi_{mix}(x, y))$ .  $\mu_{neg}^i > 0$  and  $\mu_{mix} > 0$  are the weighted coefficients. The 3D body pose estimation aims to obtain the optimal joints rotation matrix from template I-model which is determined by  $\psi = [\Delta\alpha_{\eta_1} \Delta\alpha_{\eta_2} \dots \Delta\alpha_{\eta_J}]$ . This is achieved by calculating the minimal energy cost

$$\psi' = \arg \min_{\psi=[\Delta\alpha_{\eta_1} \Delta\alpha_{\eta_2} \dots \Delta\alpha_{\eta_J}]} E_t. \quad (11)$$

Fig. 6 illustrates the construction of the energy cost function which includes two parts. One is denoted by  $E_{neg}^i$  which indicates how close the two silhouettes agree with each other in global space. If the silhouettes fit each other well, this indicates the largest level set overlap region. The level set data inside the gait contour regions has negative value and we use  $-E_{neg}^i$  to indicate the sum. We also segment the mixed level set into three parts in order to overcome the sub-optimal decisions when significant part of the body is lost. The legs usually have the higher weighted coefficient because they are associated with the most important body motion for

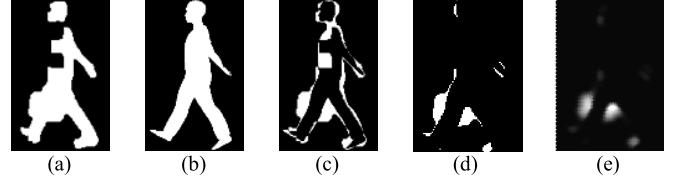


Fig. 8. preprocess of inpainting: (a) 2D incomplete gait image with a bag; (b) 2D projected gait image from 3D estimated mesh; (c) differential image of (a) and (b); (d) difference in outside contours of (b) and (c); and (e) weighted image of (d).

gait recognition. If there is a significant loss in data the leg regions would have small energy compared with other body parts and thus have make little effect or can even be ignored in the cost function. In order to address the solution being trapped in local optimum, the higher weighted coefficient is used. The other part of the cost function is denoted by  $E_{mix}$  which aids in improving the accuracy of the incomplete silhouette fitting process. Fig. 6(d) and (e) illustrate the mixture of the data set of Fig. 6(a) and two different pose projected silhouettes without weighting. The mixture of the data set is determined only by  $\phi_{mix}^{no\_weighted} = \phi_{2D}^1 \cdot S_{3D}$ . Fig. 6(d) and (e) have the same energy cost, because the incomplete data of legs are all embedded in the two different 3D projected silhouettes. However, the detail of the two estimated 3D meshes are actually different in the leg parts. Thus for optimal decisions, the silhouette weighted matrix is introduced and the  $E_{mix}$  energy is used to improve the local detail level.

*3) 2D Incomplete Gait Image Inpainting Based on 3D Body:* The first step in the 3D body reconstruction is to estimate the gait pose that is denoted by the angles of joint rotation related to the template model. The next step is to apply body shape deformation since different subjects have different body shapes. Prior to the deformation, the incomplete gait images should be repaired, i.e., the lost parts should be filled and extraneous parts, e.g., bag, hat, and ball, should be removed.

Let  $I_{2D}(x, y)$  denote the incomplete image with  $\phi_{2D}$  as its level set data.  $I_Y(x, y)$  is the 3D to 2D mapping gait image as illustrated in Fig. 5(d). It is projected from the final estimated 3D gait mesh with optimal pose parameters  $\psi_{option}$  calculated using the minimal energy cost of (11). The final estimated 3D gait mesh illustrated in Fig.5(c) is morphed from the template I-pose 3D gait model using our 3D pose estimation method. Human skeleton skinned mesh animation method is used for 3D posture morphing. The level set of  $I_Y(x, y)$  is denoted by  $\phi'_3D$ . The difference image  $I_{differ}(x, y)$  is obtained as illustrated in Fig. 8(c). The level set of the difference image is denoted by  $\phi_{differ}$ . The difference in the level sets  $\phi'_3D$  and  $\phi_{differ}$  is shown in Fig. 8(d). Let  $\phi_m1 = \phi_{differ} \cdot \phi'_3D$  and  $\phi_m2 = \phi_m1 \cdot H(\phi_m1)$ . The difference part is denoted by  $\phi_m3 = H(-\phi_m2)$ . Fig. 8(d) shows the extraneous parts are located in different parts.

To eliminate the distortion to body shape due to carried item, the level sets  $\phi_{differ}$  and  $\phi'_3D$  (respectively shown in Fig. 9(a) and (b)) are added together to give the result in Fig. 9(d) which reduces the influence of the bag. To remove the residual silhouette of the bag in edge location

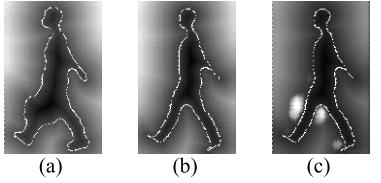


Fig. 9. Inpainting process: (a) level set of differential image with a bag; (b) level set of projected gait image from 3D estimated mesh; (c) weighted level set of (b); (d) level set adding together of (a) and (b); and (e) level set adding together of (a) and (c).

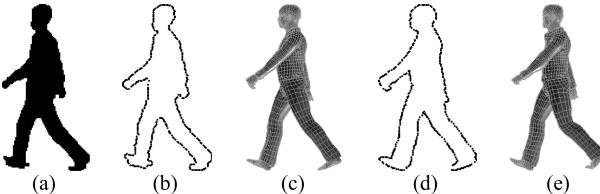


Fig. 10. Model based inpainting: (a) 2D gait silhouette; (b) markers extracted from (a); (c) pose deformed template model; (d) markers from (c); and (e) inpainting 3D model after Laplacian deformation.

we introduced the weighted matrix to enhance  $\phi'_{3D}$  in object location before adding, where the weighted matrix is

$$D'_w(x, y) = \lambda \cdot G_\sigma * d(\vec{r}, \zeta'_{3d}), \quad (12)$$

where  $\zeta'_{3d}$  is gait contour curve denoting the zero level set of  $\phi'_{3D}$ ,  $\vec{r}$  is the vectors located in  $\phi_{m3}$  and  $\lambda$  is the constant that controls the similarity between the final inpainting image and the 3D to 2D mapping gait image. The inpainting of 2D incomplete gait image is achieved by

$$\phi_{fixed} = \phi_{differ} + D'_w \cdot \phi'_{3D}, \quad (13)$$

where  $\phi_{fixed}$  and  $\phi_{differ}$  are 3D level set with coordinate  $(x, y)$  and  $z = \phi(x, y)$  is the corresponding level set value.

In Fig. 9(a) the bag is inside the contour with negative value. To eliminate the distortion, positive data is used for the sum. The bag in Fig. 9(b) and (c) have positive values. After the summation with Fig. 9(a), the bag is eliminated where the revised value is larger than zero and form the new zero level set contour. However, the revised value is sometimes not as accurate as the real gait contour illustrated in Fig. 9 (d), especially in the area close to the body contour. To address this, the larger positive value is assigned to the bag before summation as shown in Fig. 9(c). As a result, the bag is better eliminated in Fig. 9(e), the final inpainting gait image with  $\lambda = 1$ .

**4) 3D Shape Deformation Based on Complete Gait Silhouettes :** The complete gait silhouettes are used for 3D body deformation. The 3D gait pose is estimated and denoted by the angles of joint rotation related to the template model  $X$ . The model  $X$  is transformed into the estimated pose  $\psi$  by pose deformation based on model skeleton, and is denoted by  $Y = X_\psi$  as shown in Fig. 10(c). Since the instance model  $Y$  is different from the template model  $X$  in the body shape and clothing conditions, the model  $X_\psi$  needs to be processed for further shape correction by shape deformation.

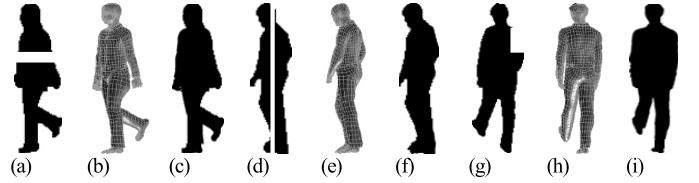


Fig. 11. 3D reconstruction using incomplete 2D gait silhouettes from different views: (a), (d) & (g) are respectively incomplete gait images at 54°, 108° and 162° gait silhouettes with manually placed bar; (b), (e) & (h) are estimated 3D gait model after pose and shape deformation; (c), (f) & (i) are 2D projected gait model from 3D mesh.

The markers  $Z = z_1, \dots, z_K$  extracted from the 2D inpainting silhouette contour are set to the target position of  $Z' = Extract_{silhouette}(X_\psi) = z'_1, \dots, z'_K$  as shown in Fig. 10(d). The shape centroid  $(x_c, y_c)$  is chosen as the reference origin. The gait contours are counterclockwise unwrapped as in [46] from the top point of the contour to convert it into a complex vector  $s = [p_1, p_2, \dots, p_N]^T$  where  $p_n = x_n + y_n j$  and  $(x_n, y_n)$  denotes each boundary pixel. To eliminate the influence of spatial scale and number of points, the vector point is equally spaced by re-sampling it to normalize its size into a fixed number (360 in our experiments).

Using the extracted silhouette landmarks, the deformed points are determined by minimizing the Laplacian deformation energy [47], i.e.,

$$\arg \min(E_L) = \arg \min \sum_{i=1}^M \|L(v'_i) - T_i d_i\|^2 + \omega_i \sum_{i=1}^K \|z_i - z'_i\|^2 \quad (14)$$

where  $M$  is the number of the vertices in 3D mesh model,  $v'_i$  is deformed  $v_i$ ,  $L(v'_i)$  is the Laplacian coordinate of the vertex  $v'_i$ ,  $d_i$  is the Laplacian coordinate of the vertex  $v_i$ ,  $T_i$  is the  $3 \times 3$  matrix which transforms  $v_i$  to  $v'_i$ .  $\omega_i$  are the weights. The resulting pose and shape deformed body model is shown in Fig. 10(e).

Fig. 11 shows several results of template based 3D gait model reconstruction from different bodies and views. The 2D gait silhouettes are segmented from CASIA dataset B [48] using background subtraction. The segmented regions are smoothed using Gaussian filter and subjected to connected-component analysis involving morphological operation of dilation to remove noisy pixels and followed by erosion to fill up any small holes inside the silhouette to give a single connected region. The pixels  $(x_{sil}, y_{sil})$  containing the human silhouette are selected as the object with maximum area [49] and are binarised using 2D Otsu thresholding. The missing data are manually filled in the 2D gait silhouettes. Using our inpainting method, the remapped inpainting 2D gait images are shown in Fig 11(c), (f) and (i).

Fig. 11(b), (e) and (h) are estimated 3D gait models using the method in Section III.B. The estimated 3D gait models are morphed from the I-pose parametric template body model by posture and shape deformation. The gait posture is determined by relative rotations of joint angles between template 3D I-model and estimated 3D model, and denoted by  $\psi = [\Delta\alpha_{\eta_1} \Delta\alpha_{\eta_2} \dots \Delta\alpha_{\eta_J}]$  estimated using our level set energy cost

function. Using the joints angles, a hierarchical skeleton with rigid bones connected by joints is reconstructed and the 3D mesh model skinned. Using the method in [35], the template 3D I-model is morphed onto the given posture gait model. 3D Laplacian deformation is then applied on the morphed posture gait model to obtain individual body shape using Laplacian deformation energy. The final estimated 3D gait models are then obtained for the corresponding views.

### C. Gait Partial Similarity Matching (GPSM)

In order to get detailed mesh models of the body in various poses and wearing conditions, 2D gait images from three or more camera views are used. This is because the 2D gait images are only represented in X-Y plane and occlusion makes body information on the other side of the 3D body unknown. Self-occlusion is one significant occlusion where a body part is partially occluded by other parts. In our proposed method only one-view gait images are used for 3D gait model estimation on a worst-case scenario that exists in practical surveillance application. Thus, the accuracy of the reverse side of the 3D model is low. For an estimated 3D gait model,  $Y_P^\theta = \{Y_{PF}^\theta, Y_{PO}^\theta\}$  at  $\theta$  view, the self-occluded body parts are associated with the 2D gait views used for the estimation. Let the estimated 3D gait model be the 3D gait model comprising the non-occluded front view part  $Y_{PF}^\theta$  and the occluded part  $Y_{PO}^\theta$ .

For gait recognition, the self-occluded parts are discarded due to their low reliability, and the non-occluded  $Y_P^\theta$  are used for gait feature extraction and recognition. As a result, the 3D gait model estimated from the 2D images with the same view and pose could be compared directly. However, gait is a dynamic movement, and it is difficult to synchronize the poses in a gait cycle, or the poses might be significantly the same for different subjects. Furthermore, the gait model estimated from different views cannot be compared directly. For an example, the gait probe feature  $Y_{PF}^\theta$  cannot be matched with the gallery feature  $Y_{GF}^\beta$  with  $\beta$  view since they have different occluded and front view parts. To address this, the partial similar gait features or the gait partial ROI elements between  $Y_{PF}^\theta$  and  $Y_{GF}^\beta$ , denoted by  $G_{ROI}^{<\beta,\theta>}$  and  $Pr_{ROI}^{<\beta,\theta>}$  where  $G_{ROI}^{<\beta,\theta>}$  is the gallery ROI (extracted from gallery data between  $\beta$  and  $\theta$  views) and  $Pr_{ROI}^{<\beta,\theta>}$  is the target ROI (extracted from probe data), are extracted. The two data can be compared directly if they are in the same pose (determined by partial surface and volume matching or nonrigid 3D matching). Since gait is not a static model and the poses in a gait cycle change, direct comparison of 3D gait models or partial features are not feasible. Thus the partial mesh constructed from the 3D gait partial ROI elements is re-projected onto 2D space to form a partial gait image. The Partial Gait Energy Image (PGEI), e.g., Fig. 12, is formed by averaging all the partial gait silhouettes in a cycle.

1) *Gait Partial ROI Elements Selection:* Partial similarity matching is used to select the common view surfaces for comparison. The selected common view surfaces are ROI elements that include mesh vertices and mesh faces. Let  $\Omega$  be the partial similar features between two different views.

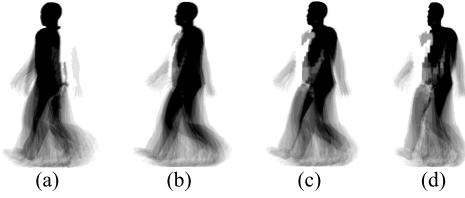


Fig. 12. Example PGEIs: (a) between  $90^\circ$  and  $126^\circ$ ; (b) between  $90^\circ$  and  $72^\circ$ ; (c) between  $90^\circ$  and  $54^\circ$ ; and (d) between  $90^\circ$  and  $36^\circ$ .

The ROI of  $M$  vertices and  $I$  mesh faces is

$$ROI = \{v_1, v_2, \dots, v_M; T_1, T_2, \dots, T_I\}, (v_{T_i}, T_i) \in \Omega, \quad (15)$$

where  $T_i = \{t_i^k | k = 0, 1, \dots, 3\} \in R^{1 \times 4}$  and  $(v_{T_i}, T_i)$  denotes a quadrilateral mesh face composed by four vertices indexed by  $t_i^k \in R$ . Let the 3D mesh estimated from  $\theta$  and  $\beta$  views be respectively denoted by  $Y_P^\theta = \{V_P^\theta, T_P^\theta\}$  and  $Y_G^\beta = \{V_G^\beta, T_G^\beta\}$ , where  $Y_P^\theta$  is the probe model and  $Y_G^\beta$  is the gallery model. To select gait partial ROI elements is to choose the non-occluded common view surface. This is achieved in two steps.

Step 1) Construct an initial ROI surface for each 3D mesh. The probe 3D body model at  $\theta$  view denoted by  $Y_P^\theta = \{Y_{PF}^\theta, Y_{PO}^\theta\}$  separates the mesh vertices and faces into two different sets. The initial ROI surface is the front (non-occluded) view  $Y_{PF}^\theta$ . To get the non-occluded surface, the occluded part is first marked out and segmented from the 3D body mesh. Since the self-occluded vertices are behind the front surface, searching all the data and faces hidden in the rear aids in obtaining the self-occluded surface. Four vertices in one mesh face are used to construct a data set  $A_s = \{v_i | i = 0, 1, \dots, 3\} \in R^{3 \times 4}$ .  $A_s$  is projected onto the X-Y plane to get a closed polygon  $D_s$ . By searching all the vertices whose X-Y plane projection belongs to polygon  $D_s$ , a larger set of vertices is constructed and denoted by  $B_s \in R^{3 \times J}$ , where  $J$  is the number of data. The mesh surface comprising the vertices in  $A_s$  is fitted with a cubic surface  $z = f(x, y) = \sum_{i=0}^3 \sum_{j=0}^3 a_{i,j} x^i y^j$  [50], where  $i$  and  $j$  define the power,  $(x, y)$  is the coordinates of the curve surface and  $a_{i,j}$  is the coefficient which is estimated from the given vertices. Each vertex in  $B_s$  is marked by  $I_{mark} = f(x, y) - z$ . The vertices belong to the occluded surface if  $I_{mark} < 0$ . By iterating through all faces in the 3D mesh, the occlusion vertices are marked out.

Fig. 13(b) shows the results of removing marked occluded vertices. Holes are found in the corresponding surface because all the faces associated with the occluded vertices are also discarded. To minimize these, the vertices in the same mesh face should be checked before removing them. They are discarded only if four of them are marked as occluded. Another problem is that part of some of the occluded surfaces might remain if any of the vertices in the retained surface belong to the occluded part. When the surface is rotated to another view, the removed surface data (shown in Fig. 13 (d)) will affect the extraction of partial similar features. To address this problem, mesh face deformation is used to transform the occluded vertices to optimized positions that result in a face with the smallest hole and overlapped areas. Fig. 14 shows

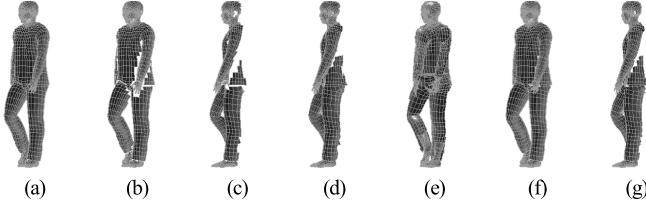


Fig. 13. Determining self-occluded vertices: (a) complete 3D gait mesh at 45° view; (b) ROI surface obtained by deleting all self-occluded vertices at 45° view; (c) ROI surface at 90° view of (b); (d) ROI surface at 90° view obtained by deleting the faces with fully occluded vertices in (a); (e) self-occluded surface obtained by our method; (f) ROI surface obtained by our method at 45° view; and (g) ROI surface at 90° view of (f).

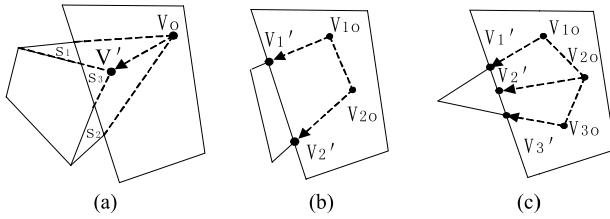


Fig. 14. Three cases of mesh face structure that should be deformed: (a) one vertex under front view surface; (b) two vertices under front view surface; and (c) three vertices under front view surface.

the three cases where a mesh face should be deformed to get high level ROI.

The first case (Fig. 14(a)) has one vertex in the mesh face under the front view surface. The occluded vertex is denoted by  $v_o = \{x_{o,1}, y_{o,1}, z_{o,1}\}$ . The vertex is transformed to the new position that makes the triangular holes (illustrated by  $s_1$  and  $s_2$  in Fig. 14 (a)) and the overlap triangle (illustrated by  $s_3$  in Fig. 14 (a)) minimum. The new location of  $v_o$  is denoted by  $v'$ . The other constraint is that the new mesh face updated by  $v'$  should have the same normal vector  $\bar{n}'$  as the former mesh face. The optimization problem is described by

$$\arg \min_{v'_o} \sum_{i=1}^3 s_i + \|\bar{n}' - \bar{n}\|^2. \quad (16)$$

Fig. 14 (b) and (c) show the other two cases respectively with two and three vertices in the mesh face belonging to the occluded surface. Let the occluded vertices be denoted by  $v_o^n$ , where  $n \in 1, 2, 3$ . The vertices in the occluded mesh face and the corresponding face above them are projected onto the X-Y plane to get two closed polygons  $D_o$  and  $D_t$ . Let  $l$  denotes the curve that intersects the two closed polygons. The newly transformed occluded vertices are denoted by  $v'_n$  which are located within curve  $l$  and the updated face retains the same normal vector  $\bar{n}'$  that satisfies  $\arg \min_{v'_n} \|\bar{n}' - \bar{n}\|^2$ .

Step 2) Obtain the partial ROI surface between  $\beta$  and  $\theta$  using the initial ROI surfaces of the front view part denoted by  $Y_{PF}^\theta$  and  $Y_{GF}^\beta$ . Let  $R_{\alpha \rightarrow \beta}$  denote the 3D rotation transformation matrix from view  $\alpha$  to view  $\beta$ . The initial ROI surfaces are affine transformed by  $Y_{PF}^\theta \cdot R_{\theta \rightarrow \beta}$  and  $Y_{GF}^\beta \cdot R_{\beta \rightarrow \theta}$ . They are transformed to each other's view. Step 1 is repeated on the two surfaces to remove the new occluded vertices. Finally, transform the resulting partial similar surface to the

same view. The gait partial ROI elements are then denoted by  $Ga_{ROI}^{<\beta, \theta>} (Y_G^\beta)$  and  $Pr_{ROI}^{<\beta, \theta>} (Y_P^\theta)$  that set  $Y_G^\beta$  and  $Y_P^\theta$  respectively as the gallery ROI and probe ROI.

2) *Partial Gait Energy Image (PGEI)*: An incomplete binary silhouettes is the projection of the common view surface and is composed of gait partial ROI elements denoted by  $Ga_{ROI}^{<\beta, \theta>}$  and  $Pr_{ROI}^{<\beta, \theta>}$ . By using the partial ROI elements, the partial similar surface meshes can be represented. Transforming them to the same gallery  $\beta$  view, and projecting to X-Y space form 2D partial images denoted by  $I_{Ga,t}^{<\beta, \theta>} (x, y)$  and  $I_{Pr,t}^{<\beta, \theta>} (x, y)$ . The partial similarity gait energy images is

$$PGEI_{<\beta, \theta>} (x, y) = \begin{cases} \frac{1}{N} \sum_{t=1}^N I_{Ga,t}^{<\beta, \theta>} (x, y), & \text{Gallery ROI} \\ \frac{1}{M} \sum_{t=1}^M I_{Pr,t}^{<\beta, \theta>} (x, y), & \text{Probe ROI} \end{cases} \quad (17)$$

where  $\beta$  is the gallery view and  $\theta$  the partial destination view.

#### IV. ARBITRARY-VIEW GAIT RECOGNITION BASED ON MULTI-LINEAR SUBSPACE ANALYSIS

To realize arbitrary-view gait recognition with partial similarity matching, a synthetic database with destination views is constructed based on GPSM. Let  $\beta_0 = \beta_G$  be the gallery view and  $\beta_{N+1} = \beta_p$  be the probe view. Let  $\varphi \in \{\beta_1, \dots, \beta_n, \dots, \beta_N\}$  be the synthetic destination views. The training partial similarity matching features are denoted by

$$D_\varphi^{\beta_G} = \left[ \begin{array}{c} Y^1(\beta_0, \beta_G) \dots Y^m(\beta_0, \beta_G) \dots Y^M(\beta_0, \beta_G) \\ Y^1(\beta_\varphi, \beta_G) \dots Y^m(\beta_\varphi, \beta_G) \dots Y^M(\beta_\varphi, \beta_G) \\ Y^1(\beta_{N+1}, \beta_G) \dots Y^m(\beta_{N+1}, \beta_G) \dots Y^M(\beta_{N+1}, \beta_G) \end{array} \right] \quad (18)$$

where  $Y^m(\beta_\varphi, \beta_G) = PGEI_{<\beta_\varphi, \beta_G>}^m$ , and  $M$  denotes the number of classification. Let  $D_\varphi = \{D_\varphi^{\beta_G}\}$ , the synthetic database  $D_\varphi$  which is represented by a higher order tensor. Multi-linear subspace analysis has been shown to be good for analysing an ensemble of images [51]. We represent the gait PGEI features of different objects with multiple partial views as a higher-order tensor  $\mathcal{D}$ . The row of the tensor denotes the PGEIs with the same partial view. The column of the tensor defines the PGEIs of the same objects. The longitudinal of the tensor illustrates the PGEIs of the different gallery views.

Tensors are multi-linear mappings over a set of vector spaces. The order of tensor  $\mathcal{A} \in R^{I_1 \times \dots \times I_n \times \dots \times I_N}$  is  $N$  elements of  $\mathcal{A}$  are denoted as  $\mathcal{A}_{i_1 \times \dots \times i_n \times \dots \times i_N}$ , where  $1 \leq i_n \leq I_n$ . The tensor  $\mathcal{D}$  is decomposed into different constituent factors by subjecting it to a generalization of N-mode single value decomposition that orthogonalizes the  $N$  spaces and decomposes the tensor as the mode-n product of N-orthogonal spaces. Thus a tensor can be expressed as a multi-linear model of factors as [51]

$$\mathcal{D} = \mathcal{Z} \times_1 U_1 \times_2 U_2 \times_3 \dots \times_m U_n \times_N U_N. \quad (19)$$

The core tensor governs the interaction between the mode matrices  $U_1 \dots U_N$ . The mode matrix  $U_N$  contains the orthonormal vectors spanning the column space of matrix  $\mathcal{D}(n)$  resulting from the mode-n flattening of  $\mathcal{D}$ .

The multi-linear analysis of different objects with multiple partial similarity gait PGEI feature ensembles leads to the tensor gait representation. The ensemble of acquired gait features is organized as a 4th-order tensor  $\mathcal{D}$ , subject, gallery views, partial destination views and PGEI pixels. The N-mode SVD algorithm is applied to compute four orthonormal mode matrices, and the tensor is decomposed as

$$\mathcal{D} = \mathcal{Z} \times_1 U_{\text{subject}} \times_2 U_{\text{g\_views}} \times_3 U_{\text{d\_views}} \times_4 U_{\text{pixels}}. \quad (20)$$

$\mathcal{Z} = \mathcal{D} \times_1 U_{\text{subject}}^T \times_2 U_{\text{g\_views}}^T \times_3 U_{\text{d\_views}}^T \times_4 U_{\text{pixels}}^T$  is the core tensor. The column vectors of  $U_{\text{subject}}$  denoted by  $c_m$  span the identification space, while their rows encode the gallery views, partial destination views and pixels invariant representation. Let the tensor base be denoted by  $\mathcal{B} = \mathcal{Z} \times_2 U_{\text{g\_views}} \times_3 U_{\text{d\_views}} \times_4 U_{\text{pixels}}$ . If the gallery gait view, partial destination view and pixels are known, the training gait features of the  $m$ th subject in gallery view  $\beta$  and partial destination view  $\theta$  can be denoted by

$$y_{<\beta,\theta>}^m = \mathcal{B}_{<\beta,\theta>, \text{subject}}^T c_m, \quad (21)$$

where

$$c_m = (\mathcal{B}_{<\beta,\theta>, \text{subject}}^T)^{-1} y_{<\beta,\theta>}^m. \quad (22)$$

The operator factor for identity is

$$P_{<\theta,\beta>, \text{subject}} = (\mathcal{B}_{<\beta,\theta>, \text{subject}}^T)^{-1}. \quad (23)$$

Given a probe gait PGEI feature  $y$  between  $\theta$  partial destination view and  $\beta$  gallery view, the identity coefficient is

$$c_{<\beta,\theta>} = P_{<\beta,\theta>, \text{subject}} y = (\mathcal{B}_{<\beta,\theta>, \text{subject}}^T)^{-1} y. \quad (24)$$

The maximum probability information of identity is selected to achieve gait recognition as

$$p(m|c_{<\beta,\theta>}) \propto \exp[-\|c_{<\beta,\theta>} - c_m\|^2 / (2\sigma^2)]. \quad (25)$$

The weighting coefficient is then used to obtain the final recognition result via majority voting, i.e.,

$$\text{Identity(probe)} = \arg \max_{m \in M} \sum_{\beta} \varpi_{\beta} p[C(c_{<\beta,\theta>}) = m], \quad (26)$$

where  $M$  is the set of classification labels and  $\beta \in \beta_G$ .  $\text{Identity(probe)}$  is the final label that is fused by the multiple classifiers.  $C(c_{<\beta,\theta>}) = m$  means the class label  $m$  learned by the individual classifier of k-nearest neighbours  $C(\cdot)$  using identity coefficient in the  $\beta$  gallery view, and  $p[C(c_{<\beta,\theta>}) = m]$  is the probability defined in (25).  $\varpi_{\beta}$  is a weighting coefficient for recognition results using different gallery dataset and it is in inverse ratio to the difference between  $\beta$  and  $\theta$ .

## V. EXPERIMENTS

### A. Experiments on CMU MoBo Database

The CMU MoBo database [52] consists of six image sequences of 25 subjects (23 males, 2 females) walking on a treadmill. Each image has a resolution of 640×480. Each subject performed 4 types of walk: slow walk, fast walk,

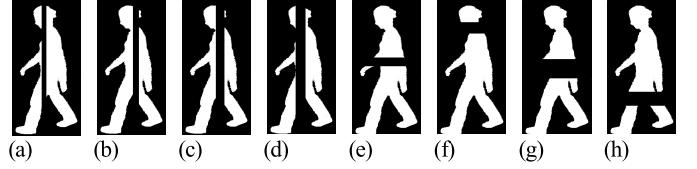


Fig. 15. Synthesized bar occlusions: (a)–(d) Vertical bars with width of 20–50 pixels; and (e)–(h) Horizontal bars with width of 40–100 pixels.

TABLE I  
RANK-1 RECOGNITION RATES WITH HORIZONTAL BAR OCCLUSIONS

Method	Horizontal bar width			
	40	60	80	100
IDTW[53]	64%	60.2%	62.4%	63.2%
GEI[8]	79.6%	80.6%	81%	79.6%
GHI[54]	54.4%	54.4%	57.8%	53.4%
GMI[55]	46.0%	46.4%	46.4%	39.6%
FD[7]	79.5%	81.4%	80.3%	80.3%
GPSM	92.6%	94.2%	93.4%	93.8%

TABLE II  
RANK-1 RECOGNITION RATES WITH VERTICAL BAR OCCLUSIONS

Method	Vertical bar width			
	20	30	40	50
IDTW[53]	66.2%	67.3%	65.8%	66.4%
GEI[8]	81%	82%	82%	80.6%
GHI[54]	52.8%	54.6%	56%	56.2%
GMI[55]	48.8%	50.8%	46.4%	48.4%
FD[7]	83.4%	83.2%	82.2%	81.4%
GPSM	94.6%	93.3%	93.1%	92.8%

inclined walk, and slow walk holding a ball. Each sequence is 11 seconds long and recorded at 30 frames per second.

In order to illustrate the advantages of our proposed method on incomplete binary gait silhouettes, occlusion is simulated by adding horizontal or vertical bar to the gallery silhouettes as illustrated in Fig. 15. A horizontal or vertical bar is added to gait silhouettes with the probability varying from 10% to 100% as in [7]. The horizontal bar width varies from 40 to 100 pixels with step size of 20 pixels, and the vertical bar width changes from 20 to 50 pixels with step size of 10 pixels.

The position of the added bar is uniformly distributed within the silhouette height or width. For comparison, the sequences captured by the frontal-view camera are chosen for the experiment. We used the fast walk sequences as the gallery and slow walk sequences as the probe set as in [7]. The test view is the same with the corresponding training view.

Table I and Table II clearly show that our GPSM outperforms the other methods in dealing with incomplete gait silhouettes. Our 2D incomplete gait image inpainting method repairs the gait images rather than mitigating the effect of imperfect silhouettes in covariate conditions that affect gait feature extraction.

TABLE III  
TWELVE EXPERIMENTS ON CMU MOBO GAIT DATASET  
(IN LATERAL VIEW)

Experiment	Gallery set	Probe set
A	Slow walk	Fast walk
B	Slow walk	Ball carrying walk
C	Slow walk	Inclined walk
D	Fast walk	Slow walk
E	Fast walk	Ball carrying walk
F	Fast walk	Inclined walk
G	Inclined walk	Slow walk
H	Inclined walk	Fast walk
I	Inclined walk	Ball carrying walk
J	Ball carrying walk	Slow walk
K	Ball carrying walk	Fast walk
L	Ball carrying walk	Inclined walk

TABLE IV  
RECOGNITION RESULTS ON MOBO DATA SET

Exp.	FSVB	WBP	STM-SPP	SGRVDL	Method[20]	PEI	GPSM
A	82%	92%	94%	96%	92%	100%	100%
B	77%	73%	93%	87%	-	92%	94%
C	-	-	-	-	-	60%	92%
D	80%	92%	91%	92%	92%	88%	96%
E	61%	61%	84%	88%	-	60%	93%
F	-	-	-	-	-	72%	94%
G	-	-	-	-	-	76%	93%
H	-	-	-	-	-	80%	94%
I	-	-	-	-	-	48%	91%
J	89%	75%	82%	87%	-	92%	93%
K	73%	63%	82%	88%	-	84%	91%
L	-	-	-	-	-	76%	92%

Twelve additional experiments in [10] were performed as shown in Table III. Given the gallery or probe sequences, GPSM is used to get 2D inpainting gait silhouettes. In this step, the object that corresponds to the carried item is eliminated and the missing data in the body repaired. As a result the gait silhouettes used for training or testing are quite accurate. Once the gait images are repaired, the GPSM features are extracted for classification. Since the gallery view and the probe view are the same, the existing methods show high recognition rates when gallery and probe sets are either the same or have small shape variation (train with S and test with S, or train with B and test with B) [43]. In order to evaluate our proposed method for robustness, experiments are chosen with gallery and probe sets under various conditions.

The methods FSVB [56], STM-SPP [57], WBP [58], SGRVDL [59] and PEI [10] evaluated under varied challenging conditions using CMU MoBo database are chosen for performance comparison. Table IV shows that our GPSM outperforms the other methods especially for ball carrying condition and inclined walk. Results of the experiments that are not presented in the original papers have been left blank

in the table. The existing methods show high recognition results when the gait silhouettes have small shape changes (e.g., slow walk vs. fast walk, and fast walk vs. slow scenarios). However, most methods are not robust to appearance changes (e.g., fast walk vs. ball carrying walk, and incline walk vs. ball carrying walk scenarios). The 2D gait silhouettes could be influenced easily by various carrying conditions. In contrast, the performance of GPSM shows satisfactory classification results across all types of gallery/probe conditions.

There are several reasons why our GPSM achieves significantly better performance. First, our estimated 3D models are embedded with skeleton structure. The motion features of the gait can be implicitly represented in our PGEIs that are robust to various challenging factors. Second, our 3D gait models are based on parametric body template with the same mesh vertices and faces. Our level set based inpainting makes our method more efficient in dealing with incomplete gait silhouettes and carrying conditions. Third, 3D parametric model is less sensitive to inclined silhouettes. This is because the initial pose of 3D models can be manipulated around X-Y-Z axes to fit any view changes whereas the 2D silhouettes cannot. As initial pose of the skeleton are included in the motion features  $\psi$ , the inclined conditions are considered in 3D pose estimation.

#### B. Experiments on CASIA B Dataset

CASIA Dataset B is a multi-view gait dataset comprising 124 subjects, and the gait data was captured from 11 views in the range  $[0^\circ \text{--} 180^\circ]$  with an interval of  $18^\circ$ . Three variations, namely in view angle, clothing and carrying condition are separately considered. There are 10 video sequences for each view of a subject: six sequences for normal walking, i.e., without wearing a coat or carrying a bag; two sequences for walking wearing a coat; and two sequences for walking with either a knapsack, a satchel or a handbag [48]. The video sequences are recorded indoor at a rate of 25 frames per second and the resolution of each frame is  $320 \times 240$ .

1) *One Gallery View Under Normal Conditions:* The performance of our view-invariant gait recognition based on partial similarity matching is evaluated in this section. First, 100 subjects are chosen randomly from the CASIA Dataset B, and the normal walk set of the selected subjects is divided into two groups. Each group contains 3 normal sequences from each of the multiple views. One group is used as gallery for training and the remaining as probe for testing. Only one gallery view is used for training.

In order to evaluate the robustness of our partial similarity matching under multi-views, we compare it with GEI-SVD [60] and GFI-CCA [9]. Fig. 16 shows the rank-1 recognition rates for views from  $0^\circ$  to  $180^\circ$ . It is clear that our method performs best. Unlike the VTM-based methods which transform the feature of probe gait data to gallery viewing angle in order to match one of the views in gallery gait dataset, our method extracts the partial similar features between two different views directly from an estimated 3D parametric gait model. It avoids the feature mismatched that often occurs in view transformation, especially with large view

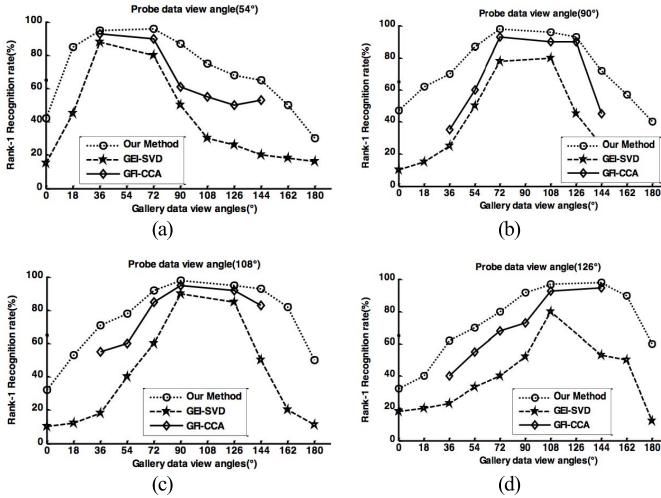


Fig. 16. Rank-1 recognition rates of different methods.

TABLE V  
RANK-1 RECOGNITION RATES ON CASIA B GAIT DATASET

Probe/ Gallery view	54°/36°	90°/108°	126°/144°
Our Method	Bag	94.2%	92.3%
	Coat	93.5%	92%
Method [28]	Bag	76.4%	73.7%
	Coat	87.9%	91.1%
RLTDA	Bag	80.8%	76.5%
	Coat	69.4%	72.1%
Robust VTM	Bag	40.7%	58.2%
	Coat	35.4%	50.3%
FT-SVD	Bag	26.5%	33.1%
	Coat	19.8%	20.6%

angle change. As a result, the VTM-based methods require a larger number of training samples to construct a more generic VTM, and better performance is achieved by learning from more varieties of gait samples. Furthermore our method which is based on parametric 3D human body is robust to incomplete gait silhouettes that aids in recognition.

2) *Under Various Conditions* : In order to evaluate the robustness of our method in various conditions under multi-views, we compare our method with VI-MGR [6], GFI-CCA [9], RLTDA [61], Robust VTM [62], FT-SVD [21] using CASIA Dataset B.

Table V shows the rank-1 recognition rates on the dataset at 54°, 90° and 126° views. The gallery gait data is under viewing angle from 36° to 144°. The known probe gait feature from one view angle is transformed into another gallery view using VTM for matching. Tables VI to VIII show our method outperforms all the existing methods, especially in bag and clothes conditions with large angle changes. Our method is robust and less sensitive to various carrying conditions including wearing a coat and carrying a bag. The parametric 3D model of the human gait is useful for view-invariant gait recognition by using prior knowledge of human body as

TABLE VI

RANK-1 VIEW-INVARIANT GAIT RECOGNITION (%) UNDER VARIOUS CONDITIONS: PROBE DATA VIEWING ANGLE 54°

Methods	Gallery	36°	54°	72°	90°	108°	126°	144°
ROI-SR	Bag	92.2	96.4	86.3	60.1	43.6	35.7	23.5
	Coat	92.9	96.1	90.5	61.8	45.4	34.8	24.2
GFI-CCA	Bag	89.3	94.2	83.5	51.6	43.9	42.2	43.5
	Coat	87.4	92.1	82.7	48.3	42.5	39.3	40.1
UMSLD-CCA	Bag	93.6	97.5	92.0	66.3	55.4	52.0	49.5
	Coat	93.1	97.0	91.4	63.5	51.2	47.6	44.3
GPSM	Bag	94.2	97.3	92.3	83.7	68.4	62.7	58.6
	Coat	93.5	97.2	91.8	82.2	65.8	60.1	51.4

TABLE VII

RANK-1 VIEW-INVARIANT GAIT RECOGNITION (%) UNDER VARIOUS CONDITIONS: PROBE DATA VIEWING ANGLE 90°

Methods	Gallery	36°	54°	72°	90°	108°	126°	144°
ROI-SR	Bag	28.4	49.1	86.2	92.9	86.6	61.2	27.2
	Coat	24.9	51.4	88.9	93.8	90.4	59.3	28.3
GFI-CCA	Bag	33.1	52.0	85.4	94.5	89.3	81.5	36.2
	Coat	30.6	52.4	83.1	92.8	85.5	77.3	34.3
UMSLD-CCA	Bag	44.2	61.7	90.4	96.0	91.9	86.3	47.9
	Coat	41.0	59.3	89.4	94.4	91.6	84.2	46.1
GPSM	Bag	62.5	82.7	93.2	96.3	92.3	88.2	67.2
	Coat	60.3	80.4	91.9	95.2	92.0	87.8	64.1

TABLE VIII

RANK-1 VIEW-INVARIANT GAIT RECOGNITION (%) UNDER VARIOUS CONDITIONS: PROBE DATA VIEWING ANGLE 126°

Methods	Gallery	36°	54°	72°	90°	108°	126°	144°
ROI-SR	Bag	26.1	37.3	45.4	68.2	92.1	94.8	91.6
	Coat	27.5	39.2	46.3	66.9	94.0	95.5	92.4
GFI-CCA	Bag	34.5	46.7	59.4	65.1	93.2	95.0	90.4
	Coat	33.4	44.6	56.1	68.3	93.4	96.5	91.2
UMSLD-CCA	Bag	42.1	52.0	62.5	70.3	94.4	96.9	95.1
	Coat	40.3	50.4	62.7	71.1	96.5	97.3	93.8
GPSM	Bag	53.2	61.9	73.2	86.5	95.6	97.1	95.1
	Coat	51.8	60.4	72.5	83.8	95.9	96.4	94.2

constraint. The 2D gait image inpainting helps to remove the bag not belonging to the body and mitigate the effect of heavy coat.

There are several other reasons why our method achieves significantly better performance. 3D parametric gait model is used for gait partial ROI elements selection. By using the extracted partial similarity ROI features, view-invariant 2D gait images can be matched directly by PGEI using GPSM. In VTM or related view transformation methods, the gait features in different views must transform to the same view for recognition. The accuracy of the VTM model is important which can influence the recognition results directly. Only when VTM is constructed using a gallery dataset that

TABLE IX

RANK-1 RECOGNITION (%) OF GPSM ON CASIA B GAIT DATASET WITH DIFFERENT VIEWS FOR TRAINING. A: TRAINING WITH 9 VIEWS FROM  $18^\circ$  TO  $162^\circ$ ; B: TRAINING WITH 4 VIEWS INCLUDING  $36^\circ$ ,  $72^\circ$ ,  $108^\circ$  AND  $144^\circ$ ; AND C: TRAINING WITH 2 VIEWS INCLUDING  $54^\circ$  AND  $126^\circ$

Probe	Normal			Bag			Coat		
	A	B	C	A	B	C	A	B	C
$18^\circ$	94	91	85	91	88	80	90	86	81
$36^\circ$	98	98	96	95	95	90	95	95	91
$54^\circ$	99	92	95	97	91	94	97	93	96
$72^\circ$	98	98	96	96	96	91	96	96	92
$90^\circ$	99	94	87	96	92	84	95	91	84
$108^\circ$	98	98	95	95	95	91	94	94	89
$126^\circ$	98	93	96	97	94	96	96	92	95
$144^\circ$	98	98	96	93	93	87	93	93	82
$162^\circ$	93	90	83	92	88	81	92	87	80

covers sufficient walking samples including various views and challenging factors can VTM be effective for the view transformation of gaits under various walking conditions. This is why VTM constructed from mixed walking conditions gives worse performance than VTM from single walking condition. In practical applications VTM is limited by insufficient training dataset for various conditions or views. In contrast, our GPSM enables gait recognition from different views especially from arbitrary views but with limited number of gait views available in training dataset. Our method is the mixture of model-based and apparent-based gait recognition that is robust for carrying items, occlusions, segmentation problem, etc. Another reason is that our GPSM is robust to wearing conditions. The 3D template model used in our method is tight clothes fitting. When the gaits are with coat condition, the silhouettes could be larger than the normal ones. After 3D body pose estimation, the larger 2D gait silhouettes with coat on body could be slimmed down by using a bigger control coefficient  $\lambda$  in (12). It is the constant that controls the similarity between the final inpainting image and the pose fitting 3D template projected gait images. The larger  $\lambda$  value makes the inpainting gait silhouettes closer to the template model. As a result, the skeleton motion features are enhanced and the coat appearance is eliminated. The coefficient  $\lambda$  can be used to adjust the composition proportion of statistic and dynamic feature of the final PGEI.

3) *Gait Recognition With Multiple Gallery Views:* Table IX shows the rank-1 recognition rate of our method on CASIA B gait dataset with multiple gallery views for training. The multi-linear subspace analysis enables our multi-view gait recognition method to use information from multi-gallery views. We match the feature of probe PGEI to the features of gallery PGEIs from all possible views, and using weighted majority voting for the final decision.

The table shows clearly that better performance can be achieved by incorporating the information from more viewing angles. A larger number of training samples with various carrying conditions and view angles will contribute to a more

accurate partial similar feature extraction, leading to even better performance in arbitrary view gait recognition under various conditions.

### C. Experiments on AVAMVG and KY4D Databases

The database AVAMVG [30] comprises 20 subjects performing 9 walking trajectories defined by  $\{t_1, \dots, t_9\}$  in an indoor environment. Of these trajectories, 3 are straight  $\{t_1, \dots, t_3\}$  and 6 are curved  $\{t_4, \dots, t_9\}$ . Each trajectory is recorded on 6 colour cameras placed around a room. The video sequences have a resolution of  $640 \times 480$  pixels sampled at a rate of 25 frames per second.

KY4D (Kyushu University 4D Gait Database) [27] comprises 42 subjects with 3D visual hull models and 2D image sequences. Each subject walks along four straight and two curved trajectories captured by 16 cameras at a resolution  $1032 \times 776$  pixels. Although the database provides 3D visual hull models of the subjects, for a fair comparative evaluation of our system we reconstruct them with our 3D parametric body model.

Our 3D parametric body reconstruction method includes two steps. The first step is to estimate the 3D body pose morphing from I-pose template model using level set energy cost function. In order to fully use the 2D multi-view gait data for accurate 3D gait reconstruction, (10) is rewritten as

$$E_t = \sum_{\theta \in \Phi} \left( - \sum_{i=1}^3 \mu_{neg}^i E_{neg}^{\theta,i} + \mu_{mix}^{\theta} E_{mix}^{\theta} \right), \quad (27)$$

where  $\Phi$  is a multi-view set determined by camera numbers. By calculating the minimal energy cost in (11), the joints rotation matrix  $\psi = [\Delta\alpha_{\eta_1} \Delta\alpha_{\eta_2} \dots \Delta\alpha_{\eta_j}]$  from the template I-pose model is obtained for pose morphing. The second step is body shape deformation. Using the multi-view extracted silhouette landmarks, the Laplacian deformation energy in (14) is redefined by combining multi-view data as

$$\begin{aligned} \arg \min(E_L) = \arg \min & \sum_{i=1}^M \|L(v'_i) - T_i d_i\|^2 \\ & + \omega_i \sum_{\theta \in \Phi} \sum_{i=1}^K \|z_i^{\theta} - z_i^{\theta'}\|^2. \end{aligned} \quad (28)$$

From the reconstructed 3D parametric gait model (illustrated in Fig.17), virtual 2D images of arbitrary view are synthesized using a method similar to that in [30]. Using the synthesized virtual 2D images with different views of a subject walking on curved trajectories, the average image of one gait cycle similar to GEI is generated. It is then used for gait recognition together with motion features of skeleton.

To evaluate the robustness of our method with limited training conditions, i.e., using dataset comprising only straight trajectories for training, for recognition of multi-view gait walking with unconstrained path, the data of straight trajectories in AVAMVG and KY4D are chosen as galleries for training while the data of curved trajectories are selected as probe for testing. Tables X and XI show that our method performed well for unconstrained gait trajectories. This is

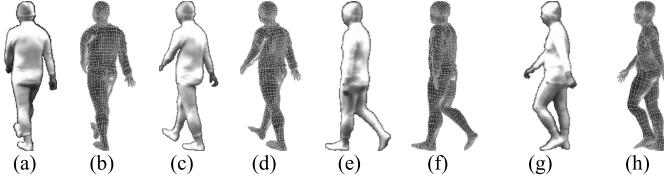


Fig. 17. Reconstruction from curved trajectories of KY4D: (a), (c), (e), (g) -2D gait images from Camera C; and (b), (d), (f), (h) the corresponding reconstructed 3D gait models.

TABLE X  
CORRECT CLASSIFICATION RATE(%) ON AVAMVG GAIT DATASET

Method	Training trajectories	$t_4$ probe	$t_7$ probe	AVG
Our method	Straight $\{t_1, t_2, t_3\}$	93.00	97.00	95.0
López [30]	Straight $\{t_1, t_2, t_3\}$	90.69	96.57	93.63
Castro [63]	Straight $\{t_1, t_2, t_3\}$	85.00	95.00	90.00
Seely [64]	Straight $\{t_1, t_2, t_3\}$	55.00	70.00	62.50
Iwashita [27]	Straight $\{t_1, t_2, t_3\}$	35.14	37.71	36.42

TABLE XI  
CORRECT CLASSIFICATION RATE(%) ON KY4D GAIT DATASET

Method	Training trajectories	$t_4$ probe	$t_7$ probe	AVG
Our method	Straight $\{t_1, t_2, t_3, t_4\}$	79.5	86.0	82.75
López [30]	Straight $\{t_1, t_2, t_3, t_4\}$	68.29	77.5	72.89
Iwashita [27]	Straight $\{t_1, t_2, t_3, t_4\}$	61.90	71.40	66.65
Castro [63]	Straight $\{t_1, t_2, t_3, t_4\}$	58.50	61.00	59.75
Seely [64]	Straight $\{t_1, t_2, t_3, t_4\}$	19.51	35.00	27.25

mainly because by using multi-view 2D gait images and our level set based inpainting process, more accurate parametric 3D body is generated from incomplete 2D data with skeleton embedded. By synthesizing arbitrary view gait images, our approach is robust to variations in view and path.

#### D. Experiments on TUM-IITKGP Database

The TUM-IITKGP database [65] contains static and dynamic occlusions with subjects walking on straight paths. It consists of 840 sequences from 35 subjects with each of the walking conditions recorded four times. The subjects were recorded in a regular walking configuration, followed by three degenerated configurations including hands in pocket, backpack and gown. The other two variations are dynamic occlusion (two subjects walking past each other) and static occlusion (two subjects just standing in the line of sight).

In our experiments, one group of sequences with regular walking is selected for training, and the remaining sequences with various conditions are for testing. The results compared with three baselines introduced in [65] are illustrated in Fig. 18. Baseline 1 (BL 1) using Colour Histogram for feature, Baseline 2 (BL 2) using GEI, and Baseline 3 (BL 3) using cropped GEI as feature for classification to eliminate the effect of backpack and gown. When walking in regular

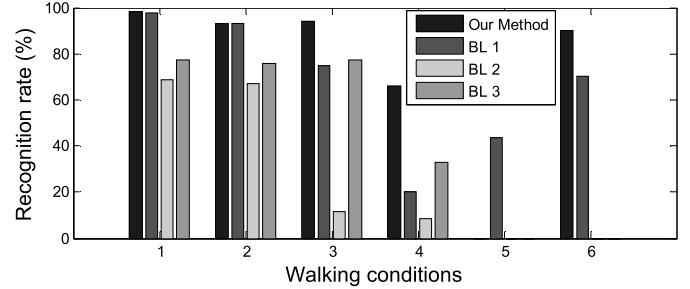


Fig. 18. Rank-1 recognition (%) on TUM-IITKGP gait dataset: 1-6 define the different walking conditions, corresponding to Regular walk, Hand in pocket, Walking with backpack, Walking with gown, Dynamic occlusion and Static occlusion respectively.

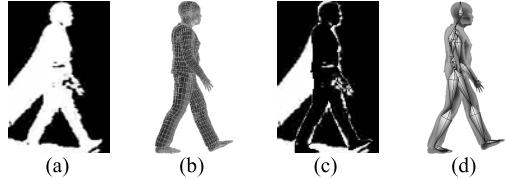


Fig. 19. Gown inpainting: (a) walking with gown; (b) estimated 3D gait model after pose and shape deformation; (c) differential image of (a) and (b); and (d) morphed skeleton of (b).

condition and with hand in pocket, the Colour Histogram Baseline method has high recognition rate. This is because it uses the colour information of the subjects that is highly correlated with the colour of the clothes. Also, it is very sensitive to lighting conditions especially for recognition under differently calibrated cameras and daytimes. However, in this experiment, the number of the subjects is limited to 35 and the subjects' clothes and illumination are unchanged. As a result, the simplest Colour Histogram Baseline method achieves good recognition result in regular walking conditions. Our approach has the highest recognition rate for walking with backpack, gown and static occlusion. This is due to our 2D incomplete gait image inpainting with weighted level set function based on 3D template body. It significantly improves the quality of incomplete 2D gait silhouettes for improved recognition as illustrated in Fig. 19.

We have not done any experiments on dynamic occlusion, because human motion tracking method should be introduced first for adequate segmentation of gait silhouettes. When the two moving subjects are fused together, it is difficult to separate them. However, the human skeleton motion information with tracking method could be used for recognition. When the moving subjects are known to be fused, we can discard the frame and use the remaining frames for recognition. These will be our future work.

## VI. CONCLUSION

In this paper, 3D parametric gait model is reconstructed from 2D video sequences using a template body model and involving pose and shape deformation. A new framework is proposed for view-invariant gait recognition using partial similarity matching. In order to extract multi-view partial similar features, GPSM is proposed. Multi-view PGEIs are synthesized to form a synthetic database for view-invariant

gait recognition. Multi-linear subspace classifier fused with majority voting is used to provide accurate identification of the subjects. The experimental results show that our proposed arbitrary view gait recognition method is robust for various conditions, e.g., occlusions, carrying items and segmentation noises. These advantages enable the proposed method to be used in many practical surveillance applications. Further work will concentrate on dynamic occlusion and other more complex variations.

## REFERENCES

- [1] I. Bouchrika, M. Goffredo, J. Carter, and M. Nixon, "On using gait in forensic biometrics," *J. Forensic Sci.*, vol. 56, no. 4, pp. 882–889, Jul. 2011.
- [2] P. Chattopadhyay, S. Sural, and J. Mukherjee, "Frontal gait recognition from incomplete sequences using RGB-D camera," *IEEE Trans. Inf. Forensics Secur.*, vol. 9, no. 11, pp. 1843–1856, Nov. 2014.
- [3] Y. Makihara *et al.*, "The OU-ISIR gait database comprising the treadmill dataset," *IPSJ Trans. Comput. Vis. Appl.*, vol. 4, pp. 53–62, Apr. 2012.
- [4] S. Sarkar, J. P. Phillips, Z. Liu, I. R. Vega, P. Grother, and K. W. Bowyer, "The humanID gait challenge problem: Data sets, performance, and analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 2, pp. 162–177, Sep. 2005.
- [5] R. Martín-Félez and T. Xiang, "Uncooperative gait recognition by learning to rank," *Pattern Recognit.*, vol. 47, no. 12, pp. 3793–3806, Dec. 2014.
- [6] S. D. Choudhury and T. Tjahjadi, "Robust view-invariant multiscale gait recognition," *Pattern Recognit.*, vol. 48, no. 3, pp. 798–811, Mar. 2015.
- [7] C. Chen, J. Liang, H. Zhao, H. Hu, and J. Tian, "Frame difference energy image for gait recognition with incomplete silhouettes," *Pattern Recognit. Lett.*, vol. 30, no. 11, pp. 977–984, Aug. 2009.
- [8] J. Han and B. Bhanu, "Individual recognition using gait energy image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 2, pp. 316–322, Feb. 2006.
- [9] K. Bashir, T. Xiang, and S. Gong, "Cross-view gait recognition using correlation strength," in *Proc. Brit. Mach. Vis. Conf., Brit.*, Sep. 2010, pp. 1–11.
- [10] A. Roy, S. Sural, and J. Mukherjee, "Gait recognition using pose kinematics and pose energy image," *Signal Process.*, vol. 92, no. 3, pp. 780–792, Mar. 2012.
- [11] H. Nakajima, I. Mitsugami, and Y. Yagi, "Depth-based gait feature representation," *IPSJ Trans. Comput. Vis. Appl.*, vol. 8, no. 4, pp. 1085–1089, Dec. 2013.
- [12] J. Tang, L. Luo, T. Tjahjadi, and Y. Gao, "2.5D multi-view gait recognition based on point cloud registration," *Sensors*, vol. 14, no. 4, pp. 6124–6143, 2014.
- [13] J. Han and B. Bhanu, "Performance prediction for individual recognition by gait," *Pattern Recognit. Lett.*, vol. 26, no. 5, pp. 615–624, Apr. 2005.
- [14] G. Zhao, G. Liu, H. Li, and M. Pietikainen, "3D gait recognition using multiple cameras," in *Proc. 7th Int. Conf. Autom. Face Gesture Recognit.*, Southampton, U.K., Apr. 2006, pp. 529–534.
- [15] G. Rogez, J. Rihan, J. J. Guerrero, and C. Orrite, "Monocular 3-D gait tracking in surveillance scenes," *IEEE Trans. Cybern.*, vol. 44, no. 6, pp. 894–909, Jun. 2014.
- [16] G. Shakhnarovich, L. Lee, and T. Darrell, "Integrated face and gait recognition from multiple views," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Kauai, HI, USA, Dec. 2001, pp. 439–446.
- [17] S. Sivapalan, D. Chen, S. Denman, S. Sridharan, and C. Fookes, "Gait energy volumes and frontal gait recognition using depth images," in *Proc. 2011 Int. Joint Conf. Biometrics*, Washington, DC, USA, Oct. 2011, pp. 1–6.
- [18] I. F. Nizami, S. Hong, H. Lee, S. Ahn, K.-A. Toh, and E. Kim, "Multi-view gait recognition fusion methodology," in *Proc. 3rd IEEE Conf. Ind. Electron. Appl.*, Singapore, Jun. 2008, pp. 2101–2105.
- [19] F. Jean, A. B. Albu, and R. Bergevin, "Towards view-invariant gait modeling: Computing view-normalized body part trajectories," *Pattern Recognit.*, vol. 42, no. 11, pp. 2936–2949, Nov. 2009.
- [20] W. Zeng and C. Wang, "View-invariant gait recognition via deterministic learning," *Neurocomputing*, vol. 175, pp. 324–335, Jan. 2016.
- [21] Y. Makihara, R. Sagawa, Y. Mukaiyawa, T. Echigo, and Y. Yagi, "Gait recognition using a view transformation model in the frequency domain," in *Proc. 9th Eur. Conf. Comput. Vis.*, Graz, Austria, May 2006, pp. 151–163.
- [22] D. Muramatsu, A. Shiraishi, Y. Makihara, M. Z. Uddin, and Y. Yagi, "Gait-based person recognition using arbitrary view transformation model," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 140–154, Jan. 2015.
- [23] W. Kusakunniran, Q. Wu, J. Zhang, and H. Li, "Gait recognition under various viewing angles based on correlated motion regression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 6, pp. 966–980, Jun. 2012.
- [24] W. Kusakunniran, Q. Wu, J. Zhang, and H. Li, "Cross-view and multi-view gait recognitions based on view transformation model using multi-layer perceptron," *Pattern Recognit. Lett.*, vol. 33, no. 7, pp. 882–889, May 2012.
- [25] M. Hu, Y. Wang, Z. Zhang, J. J. Little, and D. Huang, "View-invariant discriminative projection for multi-view gait-based human identification," *IEEE Trans. Inf. Forensics Secur.*, vol. 8, no. 12, pp. 2034–2045, Dec. 2013.
- [26] H. Hu, "Multiview gait recognition based on patch distribution features and uncorrelated multilinear sparse local discriminant canonical correlation analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 4, pp. 617–630, Apr. 2014.
- [27] Y. Iwashita, K. Ogawara, and R. Kurazume, "Identification of people walking along curved trajectories," *Pattern Recognit. Lett.*, vol. 48, pp. 60–69, Oct. 2014.
- [28] I. Rida, X. Jiang, and G. L. Marcialis, "Human body part selection by group lasso of motion for model-free gait recognition," *IEEE Signal Process. Lett.*, vol. 23, no. 1, pp. 154–158, Jan. 2016.
- [29] X. Chen and J. Xu, "Uncooperative gait recognition: Re-ranking based on sparse coding and multi-view hypergraph learning," *Pattern Recognit.*, vol. 53, pp. 116–129, May 2016.
- [30] D. López-Fernández, F. J. Madrid-Cuevas, A. Carmona-Poyato, R. Muñoz-Salinas, and R. Medina-Carnicer, "A new approach for multi-view gait recognition on unconstrained paths," *J. Vis. Commun. Image R.*, vol. 38, pp. 396–406, Jul. 2016.
- [31] A. Nandy, R. Chakraborty, and P. Chakraborty, "Cloth invariant gait recognition using pooled segmented statistical features," *Neurocomputing*, vol. 191, pp. 117–140, May 2016.
- [32] P. J. Besl and H. D. McKay, "A method for registration of 3D shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 239–256, Feb. 1992.
- [33] G. Barequet and M. Sharir, "Partial surface and volume matching in three dimensions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 9, pp. 929–948, Sep. 1997.
- [34] M. Bastioni, S. Re, and S. Misra, "Ideas and methods for modeling 3D human figures: The principal algorithms used by MakeHuman and their implementation in a new approach to parametric modeling," in *Proc. 1st ACM Bangalore Annu. Conf., Comput.*, Bengaluru, India, Jan. 2008, Art. no. 10.
- [35] B. H. Le and Z. Deng, "Robust and accurate skeletal rigging from mesh sequences," *AcM Trans. Graph.*, vol. 33, no. 4, Jul. 2014, Art. no. 84.
- [36] CMU. (2003). *Carnegie-Mellon Mocap Database*. [Online]. Available: <http://mocap.cs.cmu.edu>
- [37] D. Anguelov *et al.*, "Recovering articulated object models from 3D range data," in *Proc. 20th Conf. Uncertainty Artif. Intell.*, 2004, pp. 18–26.
- [38] S. Sivapalan, D. Chen, S. Denman, S. Sridharan, and C. Fookes, "3D ellipsoid fitting for multi-view gait recognition," in *Proc. 8th IEEE Int. Conf. Adv. Video Signal Based Surveill.*, Klagenfurt, Austria, Sep. 2011, pp. 355–360.
- [39] A. O. Balan, L. Sigal, M. J. Black, J. E. Davis, and H. W. Haussecker, "Detailed human shape and pose from images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Minneapolis, MN, USA, Jun. 2007, pp. 2861–2868.
- [40] T. Horprasert, D. Harwood, and L. S. Davis, "A statistical approach for real-time robust background subtraction and shadow detection," in *Proc. 7th IEEE Int. Conf. Comput. Vis., Frame Rate Workshop (ICCV)*, Kerkyra, Greece, Sep. 1999, pp. 1–19.
- [41] S. Osher and J. A. Sethian, "Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations," *J. Comput. Phys.*, vol. 79, no. 1, pp. 12–49, Nov. 1988.
- [42] C. Li, C. Xu, C. Gui, and M. D. Fox, "Level set evolution without re-initialization: A new variational formulation," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, San Diego, CA, USA, vol. 1, Jun. 2005, pp. 430–436.
- [43] D. A. Winter, *Biomechanics and Motor Control of Human Movement*, 3rd ed. New Jersey, NJ, USA: Wiley, 2004.
- [44] D. Cunado, M. S. Nixon, and J. N. Carter, "Automatic extraction and description of human gait models for recognition purposes," *Comput. Vis., Image Understand.*, vol. 90, no. 3, pp. 1–41, 2003.

- [45] K. Bashir, T. Xiang, and S. Gong, "Gait recognition using gait entropy image," in *Proc. 3rd Int. Conf. Crime Detection Prevention (ICDP)*, Dec. 2009, pp. 1–6.
- [46] L. Wang, T. Tan, H. Ning, and W. Hu, "Silhouette analysis-based gait recognition for human identification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1505–1518, Dec. 2003.
- [47] J. Li *et al.*, "Fitting 3D garment models onto individual human models," *Comput. Graph.*, vol. 34, no. 6, pp. 742–755, 2010.
- [48] 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. 18th Int. Conf. Pattern Recognit.*, Hong Kong, China, Sep. 2006, pp. 441–444.
- [49] R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision*. vol. 1. Reading, MA, USA: Addison-Wesley, 1992.
- [50] M. S. Floater, "Parametrization and smooth approximation of surface triangulations," *Comput. Aided Geometric Design*, vol. 14, no. 3, pp. 231–250, 1997.
- [51] Y. Chen and C.-T. Hsu, "Multilinear graph embedding: Representation and regularization for images," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 741–754, Feb. 2014.
- [52] R. Gross and J. Shi, "The CMU motion of body (MoBo) database," *Robot. Inst., Carnegie Mellon Univ.*, Pittsburgh, PA, USA, Tech. Rep. CMU-RI-TR-01-18, 2001.
- [53] S. Yu, D. Tan, K. Huang, and T. Tan, "Reducing the effect of noise on human contour in gait recognition," in *Proc. Int. Conf. Adv. Biometrics, (ICB)*, Seoul, South Korea, Aug. 2007, pp. 338–346.
- [54] J. Liu and N. Zheng, "Gait history image: A novel temporal template for gait recognition," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2007, pp. 663–666.
- [55] Q. Ma, S. Wang, D. Nie, and J. Qiu, "Recognizing humans based on gait moment image," in *Proc. 8th Acis Int. Conf. Softw. Eng., Artif. Intell., Netw., Parallel Distribut. Comput.*, Aug. 2007, pp. 606–610.
- [56] S. Lee, Y. Liu, and R. Collins, "Shape variation-based frieze pattern for robust gait recognition," in *Proc. IEEE Conf. CVPR*, Minneapolis, MN, USA, Jun. 2007, pp. 1–8.
- [57] S. D. Choudhury and T. Tjahjadi, "Silhouette-based gait recognition using Procrustes shape analysis and elliptic fourier descriptors," *Pattern Recognit.*, vol. 45, no. 9, pp. 3414–3426, Sep. 2012.
- [58] W. Kusakunniran, Q. Wu, H. Li, and J. Zhang, "Automatic gait recognition using weighted binary pattern on video," in *Proc. 6th IEEE Int. Conf. Adv. Video Signal Based Surveill.*, Genova, Italy, Sep. 2009, pp. 49–54.
- [59] W. Zeng and C. Wang, "Silhouette-based gait recognition via deterministic learning," in *Proc. 6th Int. Conf. Adv. Brain Inspired Cognit. Syst.*, Beijing, China, Jun. 2013, pp. 1–10.
- [60] W. Kusakunniran, Q. Wu, H. Li, and J. Zhang, "Multiple views gait recognition using view transformation model based on optimized gait energy image," in *Proc. IEEE Int. Conf. Comput. Vision*, Kyoto, Japan, Sep./Oct. 2009, pp. 1058–1064.
- [61] H. Hu, "Enhanced gabor feature based classification using a regularized locally tensor discriminant model for multiview gait recognition," *IEEE Trans. Circuits Syst. for Video Technol.*, vol. 23, no. 7, pp. 1274–1286, Jul. 2013.
- [62] S. Zheng, J. Zhang, K. Huang, R. He, and T. Tan, "Robust view transformation model for gait recognition," in *Proc. Int. Conf. Image Process.*, Brussels, Belgium, Sep. 2011, pp. 2073–2076.
- [63] F. M. Castro, M. J. Marin-Jimenez, and R. M. Carnicer, "Pyramidal fisher motion for multiview gait recognition," in *Proc. 22nd Int. Conf. Pattern Recognit., (ICPR)*, Stockholm, Sweden, Aug. 2014, pp. 1692–1697.
- [64] R. 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 *Proc. 2nd IEEE Int. Conf. Biometrics, Theory, Appl. Syst.*, Sep./Oct. 2008, pp. 1–6.
- [65] M. Hofmann, S. Sural, and G. Rigoll, "Gait recognition in the presence of occlusion: A new dataset and baseline algorithms," in *Proc. 19th Int. Conf. Comput. Graph., Visualizat. Comput. Vis. (WSCG)*, Plzen, Czech Republic, Feb. 2011, pp. 99–104.



**Jin Tang** received the B.Sc. degree in mechanics and the M.Sc. degree in solid mechanics from Peking University, China, in 1987 and 1990, respectively, and the Ph.D. degree in pattern recognition and intelligence system from Central South University, China, in 2002. He has been a Professor with Central South University since 2005. His research interests include computer vision and artificial intelligence.



**Jian Luo** received the B.Sc. degree in communication engineering from Hunan Normal University, China, in 2003, and the M.Sc. degree in electronic science and technology from Hunan University, China, in 2007. He is currently pursuing the Ph.D. degree in control science and engineering with Central South University, China. His research interests are pattern recognition and computer vision.



**Tardi Tjahjadi** (SM'02) received the B.Sc. degree in mechanical engineering from University College London in 1980, the M.Sc. degree in management sciences from UMIST, U.K., in 1981, and the Ph.D. degree in total technology from UMIST. He has been an Associate Professor with The University of Warwick since 2000 and a Reader since 2014. His research interests include image processing and computer vision.



**Fan Guo** received the B.S. degree in computer science and technology and the M.S. and Ph.D. degrees in computer application technology from Central South University, China, in 2005, 2008, and 2012, respectively. She is currently a Lecturer with Central South University. Her main research interests include image processing and virtual reality.