Robust Arbitrary View Gait Recognition based on Parametric 3D Human Body Reconstruction and Virtual Posture Synthesis

Jian Luo [1], Jin Tang [1,*], Tardi Tjahjadi[2], Xiaoming Xiao[1]

[1]School of Information Science and Engineering, Central South University, Changsha, Hunan 410083, China

[2]School of Engineering, University of Warwick, Gibbet Hill Road, Coventry, CV4 7AL, United Kingdom

***Abstract*:** This paper proposes an arbitrary view gait recognition method where the gait recognition is performed in 3-dimensional (3D) to be robust to variation in speed, inclined plane and clothing, and in the presence of a carried item. 3D parametric gait models in a gait period are reconstructed by an optimized 3D human pose, shape and simulated clothes estimation method using multiview gait silhouettes. The gait estimation involves morphing a new subject with constant semantic constraints using silhouette cost function as observations. Using a clothes-independent 3D parametric gait model reconstruction method, gait models of different subjects with various postures in a cycle are obtained and used as galleries to construct 3D gait dictionary. Using a carrying-items posture synthesized model, virtual gait models with different carrying-items postures are synthesized to further construct an over-complete 3D gait dictionary. A self-occlusion optimized simultaneous sparse representation model is also introduced to achieve high robustness in limited gait frames. Experimental analyses on CASIA B dataset and CMU MoBo dataset show a significant performance gain in terms of accuracy and robustness.

Keywords: Gait recognition; Human identification; Silhouette; Sparse representation.

1 Introduction

Compared with other physiological biometrics, such as fingerprint, iris and face, gait is difficult to be masked in natural body movement, whereas the appearance of a perpetrator can be disguised by wearing a hat, a pair of glasses or even by undergoing a plastic surgery. The advantages of gait recognition, i.e., recognizing subjects without their cooperation, using low-resolution video and remotely, make it particularly attractive for verification or identification purpose. Thus gait recognition has great potential, and it could be useful in various applications, such as surveillance, access control, criminal investigation and other security operations. Several evaluation systems have been developed for practical applications, e.g., gait biometrics has been used to provide forensic evidence for identification, and for criminal investigations, e.g., to provide clues for verifications of perpetrators at crime scenes [1]. In the last decade, several gait databases larger than 100 people defined as large datasets in gait field have been published, e.g., USF HumanID [2], CASIA [3], SOTON [4] and WOSG [5], and studies have shown that personal identification could be achieved where the individual and environmental factors are controlled to some extent. In recent years, larger human gait databases, e.g., OU-ISIR LP with more than 4000 people [6], have been published to evaluate the upper bound accuracy of gait recognition. It is a challenge to explore a personal identification system by gait with large population. However, gait recognition can still play a significant role in identification system, e.g., multiple biometrics could be fused to develop a more reliable personal identification system where gait recognition is used to scale down the search range for criminals from large CCTV or surveillance videos.

Gait recognition depends on the effectiveness of the extracted gait features, and the main challenges to successful recognition are covariates such as variation in view and loose clothing, speed in which the gait is performed, occlusions and carried items [7]. To address these challenges, this paper proposed an arbitrary view gait recognition system based on 3-dimensional (3D) human body reconstruction and virtual posture synthesis (AVGR-BRPS). AVGR-BRPS uses a 3D statistical parametric model of human body shapes and poses. The 3D

body shape is derived from a base mesh shape obtained by morphing, and a skeleton is embedded into the target human subject's reconstructed body model to form different postures. Body shape and poses are estimated from clothed human subjects using coarse, incomplete 3D data or 2-dimensional (2D) gait silhouettes. The gait poses with different carried items are synthesized to enable accurate identification of a subject in realistic scenarios. Since the human gait is 3D, our gait recognition method is more robust than 2D methods to variation in inclination of the ground plane, clothing and carried items by restoring the gait extracted from a video sequence into 3D space using statistical parametric model of human body shape and pose.

The two main approaches to gait recognition are model-based and appearance-based. Our method is based on a 3D gait model that is unlike a 2D skeleton model with numerous degrees of freedom [8], and the 3D voxel model [9] derived from multi-view silhouette images or point cloud scanning data. Our 3D model is created from view-invariant shape features (obtained by morphing) and pose features (described by biovision hierarchical data (BVH)). The parametric 3D gait model is more accurate than image-based visual hull model and 3D scanning voxel model due to the shape and pose deformation learned from a database of preprocessed 3D range scans of different subjects with multiple poses. Due to the skeleton embedded in our 3D human body model, virtual postures can be synthesized to form an over-complete dictionary of 3D gait dataset.

A gait energy image (GEI), extracted by averaging the silhouettes in a gait period [10], or related energy images are widely used in appearance-based approaches. The GEI reduces storage space and computation time, and can achieve high recognition rate when the gait cycle is complete. If the probe gait cycle is incomplete or it is difficult to compute the gait cycle accurately, the recognition rate decreases significantly. Since the trajectory in a gait cycle may not always be in a straight line, it is necessary to determine the view for each gait model, and adjust them to the same view in the straight line. However it is difficult to distinguish the different views in a cycle of 2D GEI or related energy images, and gait recognition rate decreases if the gaits being compared are not in the similar trajectory.

In order to address these problems, we introduce self-occlusion optimized simultaneous sparse representation model based on compressed sensing as follows. Multi-pose 3D gallery gait models constructed from a gait cycle of multiview gait silhouettes together with virtual synthesized gait models are simultaneously used for the sparse representation of probe gait model. Subject identification is achieved by combining the hypotheses of all 3D gait models and the sparse pattern. By constructing an over-complete dictionary of 3D gait and simultaneous sparse representation model, the simultaneous sparse representation vectors are reconstructed for 3D probe gait models which has been self-occlusion optimized in a gait cycle. The reconstruction error is used for classification.

The most notable advantages of AVGR-BRPS are: (1) by utilizing clothes-independent 3D human pose and shape estimation model learned from a database of parametric 3D shapes and simulated clothes, it reconstructs parametric 3D human body from 2D or 2.5D data to achieve robustness to variation in significant view changes, heavy or loose clothing variation, difference in carried items, and occlusions, e.g., missing body parts or segmentation noise; (2) by using carrying-item posture synthesized model, synthesized virtue carrying postures are used to extend the dictionary of 3D gait which makes AVGR-BRPS more robust and improves its recognition rate; (3) by constructing an over-complete dictionary of 3D gait database and using self-occlusion optimized simultaneous sparse representation model, it achieves view-invariant recognition that does not require a complete gait cycle data and the subject to walk in a similar trajectory.

The rest of this paper is organized as follows. Section 2 presents the related work. Section 3 presents AVGR-BRPS. Section 4 presents the experimental results and Section 5 concludes the paper.

2 Related work

Video sequences captured by multiple cameras have been used to create a 3D human skeleton model in [11] where static parameters and dynamic features are combined to describe gait. Although the model enables robustness to changes in viewpoints and clothing variation, the recognition method uses multiple cameras and the initial subject's pose is extracted manually. The use of recursive Bayesian sampling and pose-viewpoint manifold for view invariant 3-D gait tracking with a high perspective effect is proposed in [12]. However, its invariant monocular 3-D human pose tracking must operate in man-made environments and assumes the subject moves on a known ground plane. The method in [8] uses an articulated human model constructed from 3D volume sequences to address a viewpoint-free framework. The method in [13] uses an image-based visual hull to construct a 3D gait model. The method in [14] uses a multi-view synthesizing method based on 2.5D point cloud registration to generate 3D gait model to achieve view-invariant gait recognition. The accuracy of the above-mentioned 3D models is limited due to the use of 2D images without depth information or point cloud data captured by low-resolution 3D scanner with missing data.

Multi-view gait recognition based on view transformation model (VTM) learns projection relationships of multi-view gaits features by singular value decomposition (SVD). It realizes multi-view gait recognition by transforming all gait views onto the same view. The VTM method in [15] transforms a probe gait onto the virtue view that exists in the training database for multi-view gait recognition. The method in [16] uses sparse regression-based VTM, where regression is used to formulate and model correlated motions among gaits under different views. In [17] the VTM is constructed using multi-layer perceptron as a regression tool, and a selected region of interest (ROI) from source view(s) is used to estimate gait from target view. The method in [18] uses an arbitrary VTM that accurately matches a pair of gait traits from an arbitrary view. The gait recognition is realized by projecting the 3D gait volume sequences onto the same views of 2D gait silhouette sequences as the target views. Although VTM related multi-view gait recognition methods demonstrate significant advantage over other approaches, they 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. The parameters of the VTM are sensitive to the training multi-view images.

By extracting view-normalized or view-invariant gait features, several view-invariant gait recognition methods have been proposed**.** In [19], gait silhouette features that can represent gait dynamics and reflect the view variation are extracted, and deterministic learning theory is introduced to achieve view-invariant gait recognition. The method in [20] uses the 2D trajectories of both feet and head extracted from tracked silhouettes as gait features. It uses view normalization based on homography transformation to make the walking appears to have been observed from a fronto-parallel view. However, self-occlusion or significant changes in views significantly affect the extraction of trajectories, thus making the method less practical for realistic scenarios. The method in [21] uses view-invariant gait features derived from pose estimation of the joint positions of walking subjects. In [22] view-normalization is performed in the input gait silhouettes to normalize gaits from arbitrary views by using corresponding domain transformation obtained via invariant low-rank textures. In [23], the multi-view gait features are extracted by view-invariant discriminative projection and matched without knowing or estimating the viewing angles. However, by using linear projection, it is impossible to achieve the ideal performance when the viewing

angle changes significantly. The method in [24] uses a sparse local discriminant canonical correlation analysis (CCA) to model the correlation of gait features from cross views, and uses patch distribution feature to classify the view. Thus, it does not require the gait to be transformed to a virtue view. However, the gait features are less correlated when the gait views are significantly changed.

Numerous gait recognition methods have been proposed to deal with multiple covariates (e.g., occlusion, clothing invariant, carrying conditions, speed and incomplete sequences). The method in [25] uses a dynamic frame difference energy image to reduce the influence of incomplete silhouette caused by occlusion or missing body parts. To address substantial clothing variation, the method in [26] divides the subject's body into eight clothing related sections assigned with different weight for matching. It requires a large-scale gait dataset with clothing variations for gait training, making it impractical. The method in [27] successfully extracts wavelet based silhouettes of a subject without carried items or a subject with heavy clothing from infrared thermal images. The gait entropy image introduced in [28] is designed mainly for mitigating the effect of changes in covariate conditions that affect gait feature extraction rather than the gait itself. To address the situation where full gait cycle information is not always available, the method in [29] uses skeleton joint information and fractional gait energy image for frontal gait recognition. The method in [7] uses view-invariant multiscale gait recognition method (VI-MGR) to achieve robustness to variation in clothing and presence of a carried item. The method in [30] selects the most discriminative human body part based on group Lasso of motion to reduce the intra-class variation so as to improve the recognition performance. In [31], sparse coding multiview hypergraph learning re-ranking (SCMHLRR) method, integrated sparse coding and multi-view hypergraph learning are explored for recognizing a pedestrian under uncooperative setting, i.e., carrying factors, walking speed and noise. Most of these methods address different covariates with limited gait views, mainly using frontal view of the gallery and probe gait sequences, or the experiments only focus on one condition at a time. They hardly explore the covariate conditions simultaneously, i.e. factor of views, carrying items and clothes variation together. In fact, it is usually hard for a 2D training dataset to cover all conditions, especially different carrying and clothes variations that affect the overall body shape directly**.**

AVGR-BRPS is proposed to address the above-mentioned gait recognition problems with multiple views and robustness for different clothing and various carrying conditions. An accurate and unified I-posture 3D parametric human body model is used to estimate 3D gait model from multi-view 2D gait silhouettes using clothes-independent 3D human pose and shape estimation method. Since a gait period comprises several gait frames, it is time consuming to estimate the different posture 3D gait models from the initial standard I-posture. In order to overcome this problem, we also propose an optimized approach for 3D gait model reconstruction based on extended standard gait models in a gait period. The k-means clustering is used to determine the key standard gait postures. Using the 3D parametric gait model reconstruction method, different gait models are obtained and are used as galleries in the training process. To achieve high recognition rate, virtual gait models of postures with different carrying items conditions are synthesized to create an over-complete 3D gait dictionary, and a self-occlusion optimized simultaneous sparse representation model is introduced.

3 Proposed method: AVGR-BPRS

3.1 Overview

A general parametric 3D human model is normally created from 3D scans by combining multi-view body data of a subject with tight clothing walking past multiple scanners, where the scanners are typically expensive and limited for fast estimation of the body shape underneath the clothes. Also, the use of multiple sensors in training or recognition is often not practical. A 2.5D gait voxel model reconstructed by a range sensor like Kinect is used for gait recognition in [14] in order to overcome the high computational cost of 3D gait modelling. However, 2.5D gait model cannot address the problems of the lack of robustness to covariates such as heavy or loose clothing, differences in carried items, etc. In order to overcome these problems, we adopted a unified and clothing-independent parametric 3D modelling method for 3D gait recognition. We estimate both the body shape and posture from motion sequences of a clothed subject. The overview of AVGR-BPRS is shown in Fig. 1.

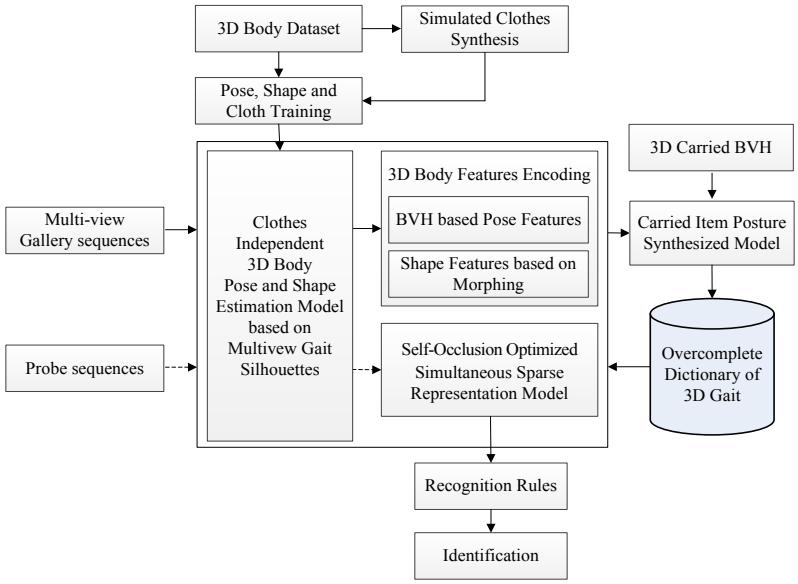


Figure 1. Overview of AVGR-BPRS.

We use the statistical model of body pose and shape like SCAPE which has been used to accurately identify a human subject with numerous challenging factors of realistic scenarios [32] to estimate 3D body pose and shape in a gait cycle using gait silhouettes. In our work, the model is re-trained on a dataset of dense full-body 3D scans, obtained from the software Makehuman which incorporates 1170 morphings for effective parametric modelling. The parametric body model in Makehuman is optimized for subdivision surfaces modelling with 15128 vertexes, which is suitable for gait modelling. We synthesized simulated clothes for all the training bodies in order to learn the clothes modelling used in the clothes-independent 3D human body reconstruction process.

Following the estimation of the 3D gait model, static shape features such as gender, body height, weight, body fat scale, body fat percentage, percentage of muscle tissue etc., and dynamic motion features like initial pose of the skeleton and joints position are encoded separately. By storing these semantic variables, the 3D gait model can be expressed simply, and by using these estimated high-level semantic variables, body shape morphing can be applied and appropriate body reconstructed as necessary. The motion capture data format of BVH encodes the dynamic motion features, such as the position of the root and orientation of joints. The skeleton structure of BVH is shown in Fig. 2. Embedding the skeleton in the body model enables virtual poses to be synthesized to form an over-complete dictionary of 3D gait dataset. The self-occlusion optimized simultaneous sparse representation model based on compressed sensing enables arbitrary view gait recognition that does not require a complete gait cycle data in similar trajectory.

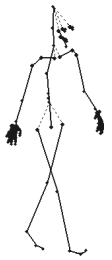


Figure 2. Skeleton structure of BVH.

3.2 3D parametric statistical model of body pose and shape

The automatic generation of realistic 3D human body according to several constraints is important in 3D body pose recognition, human tracking or 3D gait recognition. In gait applications, the arbitrary view gait recognition problem has become one of the major issues that must be overcome. The traditional multi-view gait recognition based on 2D gait silhouettes is not robust to varying practical environments. By lacking the unified and prior knowledge of human body, the recognition rate decreases dramatically when there is occlusion or missing data in a gait cycle. The carrying problem has also not been well addressed. Existing methods only focus on mitigating the effect of changes in covariate conditions that affect gait feature extraction rather than the gait itself. For an example, in order to improve the gait recognition rate for ball carrying in MoBo dataset only the lower part of the body is used. 3D gait recognition approaches have been explored to address this problem. However, the 3D gait model used is either visual hull appearance model or 3D scanning voxel model, which requires either multiple cameras or numerous multi-images. None of the models embed the human skeleton for motion feature capture or generation of virtual pose. In this paper we introduce a statistical model of human body pose and shape in our 3D gait model reconstruction process. The model learns a probability distribution from a 3D human body dataset with varying shapes and poses.

In [33] a statistical shape model is learned by principal component analysis (PCA) and used to synthesize individual subjects using a template-based method. The pose of the morphing models is similar to the template. In [32] a data-driven method is proposed for building a human shape model that spans variation in both subject shape and pose that encodes the surface relative to an embedded skeleton. In [34] a rotation invariant encoding of parametric 3D human body is used to create a model with a given posture by training semantically meaningful regressors.

Publicly available 3D human shape dataset is used in our parametric model of body pose and shape training. The database of k input scans with s as subject identifier and p as pose is denoted as $S_{s,p}$. All model $S_{s,p}$ are parameterized by the parameter vector Y consisting of n points $P_t = p_1^t, ..., p_n^t$ and l faces $T^t = t_1^t, ..., t_l^t$. All models are in full point-to-point correspondence and each model mesh has the same set of points and faces. The parameterized models are then encoded by the rotation invariant model representation approach in [34] where the differences between a pair of scans can easily be extracted in training.

First, a parameterized model with a standard posture is selected as a template. Each faces t_i in the remaining pose models is encoded as a transformation $U_i = R_i \cdot S_i$, where R_i are rotation matrices and S_i is stretching deformation. The encoding is achieved by storing relative rotations between pairs of adjacent faces, i.e., [34]

$$R_{i,j} = R_i \cdot R_j^{-1}, \tag{1}$$

where i and j are adjacent faces. In the reconstruction process, the rotation of each faces R_i is recovered from the relative rotations $R_{i,j}$ by solving the sparse linear system

$$R_{i,j} \cdot R_j - R_i = 0. \tag{2}$$

Using the rotation invariant model, body shape and pose are jointly encoded. A 3D parametric statistical model is then learned based on the ideas in [32] and [34]. The purpose of the training is to learn the pose and shape related to semantic values such as gender, body height, weight, body fat scale, percentage of muscle tissue or joint angle etc. Using a regression approach, our system morphs a new subject with some constant semantic constraints.

Denote a set of encoded parametric model $A = [a_1 \ldots a_i]$ and discretised semantic values function $l = [l_1 \ldots l_i]^T$. By solving

$$\arg\min_{[o,s]} [1 \mid A] \begin{bmatrix} o \\ s \end{bmatrix} - l \tag{3}$$

in the least-squares sense using a non-linear support vector regression, a gradient direction s and a corresponding offset o is obtained [34].

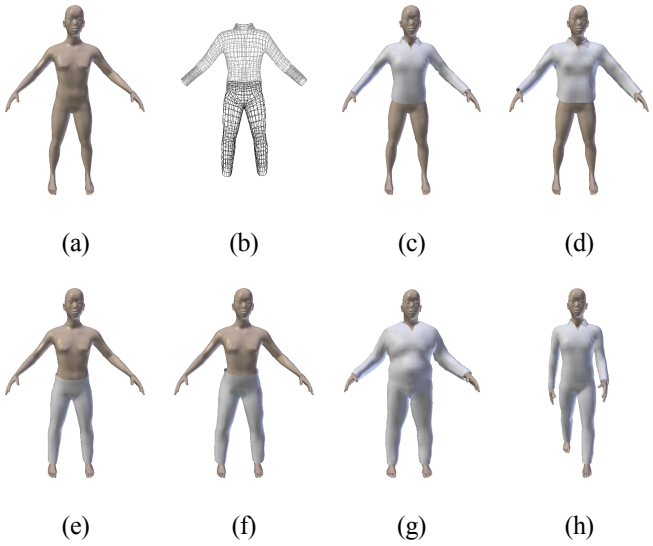
Using the learned parameters s and o, and setting discretised semantic values l as constant constraints, a new morphable model is created. The semantic values l are then separated into two sets: static body shape parameters such as body height, body weight etc. denoted by $S = [\beta_1, ..., \beta_k]$ and motion posture parameters such as joint angles denoted by $\psi = [\gamma_1, ..., \gamma_k]$.

3.3 3D parametric statistical models with clothes

The scanned human bodies used for training in publicly available 3D datasets are wearing body-fitting clothes to reduce the influence of the clothes. However, in gait recognition or other surveillance applications, the subjects are wearing different clothes, e.g., even heavy or loose clothing. As a result the estimated body shape may be inaccurate. Current solutions to this problem are learning about likely body shapes using different methods, especially on the naked body [35, 36, 37]. Existing methods perform well on scans of subjects with body-fitting clothes, but are poor in estimating body shape and postures from motion sequences of 2D images of subjects with varying types of clothes. In order to overcome the problem, simulated clothes are synthesized on human body to predict realistic human body shapes and postures as illustrated in Fig. 3. Two sets of simulated clothes are introduced on the same body of Fig. 3(a): one set is body-fitting clothes (shirt in Fig. 3(c) and trouser in Fig. 3(e)); and one set is loose clothes that is 15% larger (shirt in Fig. 3(d) and trouser in Fig.3 (f)). The clothes are deformed from the same template clothes. They are shown separately to avoid the same body with loose clothes being perceived as a larger body with tight clothes.

The simulated clothes are predicted for different body shapes parameters S using our proposed clothes model. First, a template clothes C_T is synthesized manually from our template body shape using the publicly available software Blender. The template is a parametric model similar to the human body consisting of m points and L faces $CT^t = ct_1^t, \ldots, ct_L^t$. Second, using the clothes model, we synthesized all the clothes C_i for each training human body shape using rotation and stretching deformation for each faces CT^t. The deformed faces are encoded as a transformation $CU_i = CR_i \cdot CS_i$ related to the corresponding template clothes faces. Third, the clothes model is then learned using the method similar to the 3D parametric statistical model by solving a

non-linear support vector regression problem in Eq. (3). In this situation the discretised semantic values are only the shape parameters and function $l = [l_1 \ldots l_i]^T = S$. As a result, if the body shape parameter S is known, the corresponding clothes mesh model can be predicted and fitted to the body. The *kth* body with clothes is denoted by $Y_k(S, \psi_k, C_p(S))$, where $p = 0$ indicate body-fitting clothes and $p = 1$ is loose clothes.



(a) (b) (c) (d)

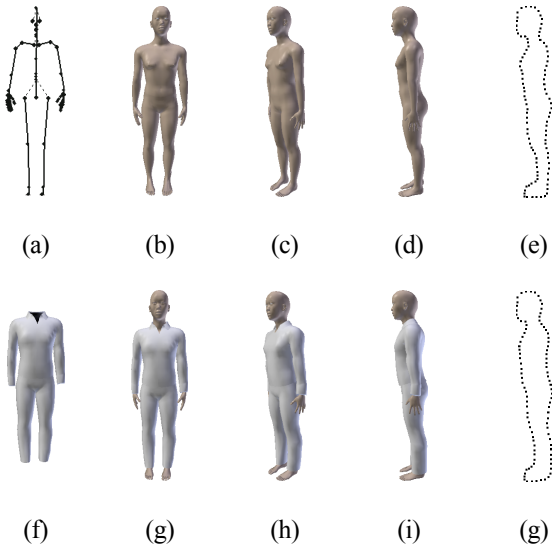
(e) (f) (g) (h)

Figure 3. 3D body template: (a) model; (b) mesh model; (c) with body-fitting clothes; (d) with loose clothes; (e) with body-fitting clothes; (f) with loose clothes; (g) fat body shape with body-fitting clothes; and (h) with clothes in walk posture.

3.4 3D pose and shape estimation underneath clothes

We use multi-view 2D or 2.5D gallery sequences for 3D pose and shape estimation based on 3D parametric model of pose and body shape with gait silhouette as constraint.

1) Skeleton based posture and shape estimation



(a) (b) (c) (d) (e)

(f) (g) (h) (i) (g)

Figure 4. 3D body template: (a) model skeleton; (b) & (g) model at 0° view; (c) & (h) model at 45° view; (d) & (i) model at 90° view; and (e) & (g) model silhouette at 90° view; and (f) model simulated clothes.

A symmetric morphable template with standard I-posture in a gait cycle is shown in Fig. 4. A skeleton structure of BVH is fitted to the template. The standard 3D model is projected onto 2D space at θ view, and denoted

by $B_\theta(S,\psi,C_p(S))$, where S is body shape parameter, ψ is joint angle of I-posture and $C_p(S)$ is clothes model parameter with S body shape and p type.

(a) (b) (c) (d) (e) (f)

Figure 5. Multiview gait silhouettes at: (a) 0° view; (b) 36° view; (c) 72° view; (d) 90° view; (e) 126° view; and (f) at 162° view.

The silhouette contour at θ view is extracted as a set of markers in 2D or 2.5D galleries as shown in Fig.5. They are denoted by

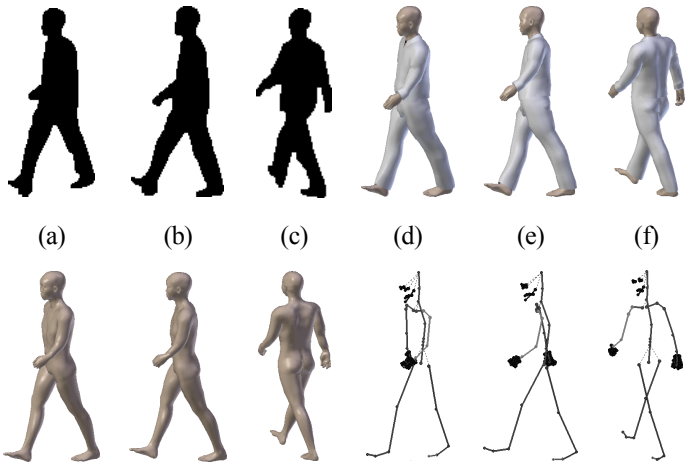
$$Z^\theta = Extract_{silhouette}(X^\theta) = z_1^\theta,...,z_I^\theta \tag{4}$$

where X^θ is 2D or 2.5D gallery at θ view, and I is the maximum discrete point number. The silhouettes are used to estimate the shape features, motion features ψ and clothes type p using template model by minimizing the silhouette cost function

$$\arg\min_{[S,\psi,p]} \sum_{\theta\in\Phi}\sum_{i=1}^{I}\left\| Extract^i_{silhouette}(\{B_\theta(\psi,S,C_p(S))\}) - z_i^\theta \right\|^2 , \tag{5}$$

where Φ is a multi-view set, and $Extract^i_{silhouette}(\cdot)$ is the ith silhouette contour marker extracted from projected template $B_\theta(S,\psi,C_p(S))$.

By solving the optimization problem in Eq. (5) the 3D pose and shape are estimated, and the parametric 3D body without clothes $Y(S,\psi)$ is reconstructed using features S and ψ. Fig. 6 shows the raw gait images and their estimated model in 2D plane.



(a) (b) (c) (d) (e) (f)

(g) (h) (i) (j) (k) (l)

Figure 6. Raw gait images and the estimated models with and without clothes in 2D plane: (a), (d) & (g) at 36°view; (b), (e) & (h) at 90°view; (c), (f) & (i) at 126°view; and (j)-(l) the corresponding BVH skeletons.

2) Encoding 3D gait features

The 3D gait features estimated in Step 1 include static shape features S and motion features ψ such as joint angles. Since a gait cycle comprises several gait models with different postures, we store all the motion features ψ in a gait cycle as BVH data. A BVH skeleton data has two parts, a header section that describes the hierarchy and initial pose of the skeleton, and a data section that contains the motion data.

3) Optimized for 3D gait Pose and Shape Estimation

Using multi-view silhouettes cost function, 3D shape and pose can be reconstructed. However, a problem is encountered if the pose is significantly different to the initial standard I-posture. In order to solve this problem, a group of standard 3D gait models with different postures in a gait cycle are constructed. In STS-DM [38], a gait period is divided into ten phases, and the silhouettes of the ten phases are manually extracted from stance phase to swing phase. Unlike STS-DM, AVGR-BPRS uses a clustering algorithm to confirm the phases. The 129 silhouettes of 5 subjects at $\theta = 90°$ view in CASIA B data set within a gait sequence are collected as observations. The discrete gait silhouette contours of the observations are first obtained using Eq. (4), where I is set to 128. The $n = 129$ observations are denoted by $\{x_1, x_2, ..., x_n\}$, where each observation is a 128-dimensional real vector. *k*-means clustering is used to partition *n* observations into *k* clusters in which each observation belongs to the cluster with the nearest silhouette mean. Setting *k* classes $c = \{c_1, c_2, ..., c_k\}$ and u_j as *jth* cluster mean, the problem reduces to minimizing the within-cluster sum of squares

$$\arg\min_{c} \sum_{j=1}^{k} \sum_{x_j \in c_j} \left\| x_j - u_j \right\|^2. \tag{6}$$

Fig. 7 shows the *k*-means clustering results in 3D space. The different symbols denote the observations belonging to different clusters, and the filled circles represent the mean of points in c_j. Since the observation data is a 128-dimensional real vector, the multi-dimension data is reduced to 3D space by multidimensional scaling (MDS). Fig. 8 shows ten mean silhouettes after K-means clustering.

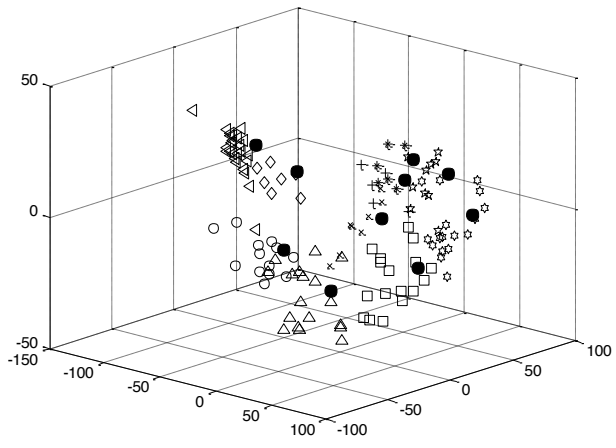


Figure 7. *k*-means clustering in 3D space by MDS.

Following *k*-means clustering, ten silhouettes $X^{\theta} = \{x_1^{\theta}, x_n^{\theta}, ..., x_k^{\theta}\}$ with the minimal distance to their within-cluster mean are selected as the template objects. CASIA dataset B [3], a large multi-view gait database, is used for 3D template model construction. There are 124 subjects, and the gait data was captured from 11 views. Three variations, namely view angle, clothing and carrying condition are separately considered. The subjects at $\theta = 90°$ are

used to estimate the ten phases, since the lateral view of gait contains the most information of gait. The corresponding gait silhouettes in ten other views associated with the same phases to the template objects are extracted to construct the multi-view silhouettes set $X = \{X^{\theta}\}$, where $\theta \in \Phi$ and $\Phi = \{0°, 18°, 36°, 54°, 72°, 90°, 108°, 126°, 144°, 162°, 180°\}$. The subjects and silhouettes of the corresponding phases in the other views are manually selected and extracted. The multi-view silhouettes set $X = \{X^{\theta}\}$ is then used to reconstruct the 3D parametric template gait models using Eq. (5).

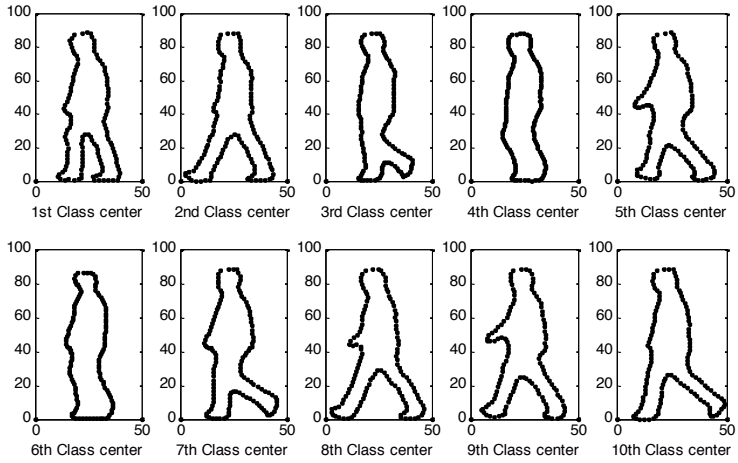


Figure 8. Ten mean silhouettes after K-means clustering.

The 3D parametric template models in a gait period with *k* phase are denoted as $Y_k(S,\psi_k)$. $B_k^{\theta}(S,\psi_k)$ are the 2D silhouettes that are projected from the 3D model Y_k at θ view. They are used for fast 3D gait pose and shape estimation using Eq. (5). To estimate the most similar *k* phase from the 3D parametric template models Y_k in a gait period of any given gait sequence, we apply silhouette comparison based on weighted Krawtchouk moments to obtain similarity scores [38], i.e.,

$$\arg\min_{n} \left\| Rf_{ROI}^{k} - Tr_{ROI}^{n} \right\|^2, \tag{7}$$

where the reference ROI, i.e., Rf_{ROI}^{k}, is the bottom segment of gait silhouette $B_k^{\theta}(S,\psi_k)$, the target ROI, i.e., Tr_{ROI}, is set to the same silhouette segments of all frames of a subject's gait period, and *n* denotes the frame number. The frame whose Tr_{ROI}^{n} results in the least squares with the corresponding Rf_{ROI}^{k} is extracted as the *kth* phase and the process continues by comparing the next Rf_{ROI}^{k} with the remaining Tr_{ROI}^{n} until all ten phases are obtained. The 3D shape feature S and posture features ψ_k of *k* phases are then estimated using the method in Section 3.3 with the corresponding template gait model Y_k. Using the estimated parameters S and ψ_k, the 3D gait models in any gait sequence are reconstructed.

3.5. Carrying-Item Posture Synthesized Model

The 3D gait models are encoded using the shape features S based on morphing which is invariant to views and the motion features based posture parameters ψ. The identification of subjects is achieved by comparing S and ψ. Unlike the view-invariant and carrying unrelated shape features S, the posture associated gait motion features ψ are much sensitive to the carried items and carrying conditions that influence the body skeleton pose, e.g., ball

carrying makes the hand in a static state which leads to significant change in silhouettes compared to the training silhouettes without carried items. Existing methods especially the model-free GEI or silhouette-based approaches just mitigate the effect of changes in carrying conditions that affect gait feature extraction. These methods cannot estimate the silhouettes that are not in the training dataset, thus reducing the recognition. To address such a problem, the carrying item posture synthesized model is used to synthesize virtue carrying postures to extend the dictionary of 3D gaits.

1) Different 3D Carrying BVH Data Extraction

Multiple 3D BVH motion data is collected according to *M* different sets of carried items. Let $A_{m,\theta}^{k,n}$ denote the *nth* subject silhouette contour vector extracted at θ view from 2D gait silhouette with *mth* carried item that corresponds to the *k*th phase in a gait period. The 3D parametric gait models are then reconstructed using silhouette contour vector $A_{m,\theta}^{k,n}$ using the method in Section 3.3. The BVH motion data of skeleton $BVH_m^{1:k,n}$ based on posture parameters $\psi_m^{k,n}$ is then extracted. The mean skeleton based BVH motion data with *mth* carried item is given by

$$BVH_m^{1:k} = \frac{1}{N}\sum_{n=1}^{N} BVH_m^{1:k,n}. \tag{8}$$

The skeletons extracted from different carrying conditions are defined as carrying item or non-standard skeletons.

2) Embedding Carried Item Skeletons to 3D Gait Models

The BVH motion data is skeleton based and it is the representation of the posture parameters. Using the parametric 3D model reconstruction, the carried items are excluded from the body. The carrying-item skeletons extracted from different carrying conditions are then embedded to the 3D gait models that represent normal walking. As a result, the virtual carrying-item gait models are synthesized. Fig. 9 illustrates the process of synthesizing the virtual gait model from the BVH skeleton data of carrying a ball. Using the proposed method, the virtual carrying-item gait models can be synthesized from 3D normal-walking models with different body shape and posture data.

The carried items mostly influence the carrying associated part of the skeleton, and the other skeleton data of the gait associated with a subject will not be changed. Therefore, the virtual synthetic skeleton comprising carrying data from carrying-item skeleton and the remaining data from the standard 3D gait skeleton is synthesized first. Since multiple carrying conditions may lead to different carrying associated parts, these parts could include different bones and joint angles of the body skeleton, and they are defined as the *ROI* of carrying item skeleton. For an example in the ball carrying condition, the *ROI* may include the bones and joints associated with the upper and lower arm of the two hands as shown in Fig. 9(f). The ROI from the carrying-item skeleton is denoted as $BVH_m^{1:k}(ROI)$. The other skeleton data from the normal 3D gait model is denoted by $BVH_{m,S}^{1:k}(ROI^{-1})$, where S is body shape parameter. As shown in Fig. 9 (h), the ROI^{-1} of normal-walking skeleton defines the bones and joints excluding the *ROI* associated ones. They are $BVH_m^{1:k,n}$ and are used to obtain the synthesized skeleton data $BVH_{m,S}^{1:k}(synthesized)$ as shown in Fig. 9(i). Using S and $\psi_{m,S}^{1:k}$ extracted from $BVH_{m,S}^{1:k}$, the virtual carrying-item posture synthetic model $Y_k(S,\psi_{m,S}^k)$ is reconstructed, where $B_k^\theta(S,\psi_{m,S}^k)$ is the 2D silhouette from the 3D model at θ view. The carrying-item posture synthesized model is given by

$$Y_k(S,\psi_{m,S}^k) = F_{3D}(S, f_{ROI}(\psi_m^k) \cup f_{ROI^{-1}}(\psi_S^k)), \tag{9}$$

where $f_{ROI}(\cdot)$ denotes the function that extracts the *ROI* data of the carrying item skeleton, and $f_{ROI^{-1}}(\cdot)$ denotes the function that extracts the rest of the skeleton. $F_{3D}(\cdot)$ is the 3D reconstruction process using S and $\psi_{m,S}^{k}$. Using the carrying-item posture synthesized model, virtual gait postures with different carrying conditions that are not included in the training data set are synthesized using the prior knowledge of carrying items BVH.

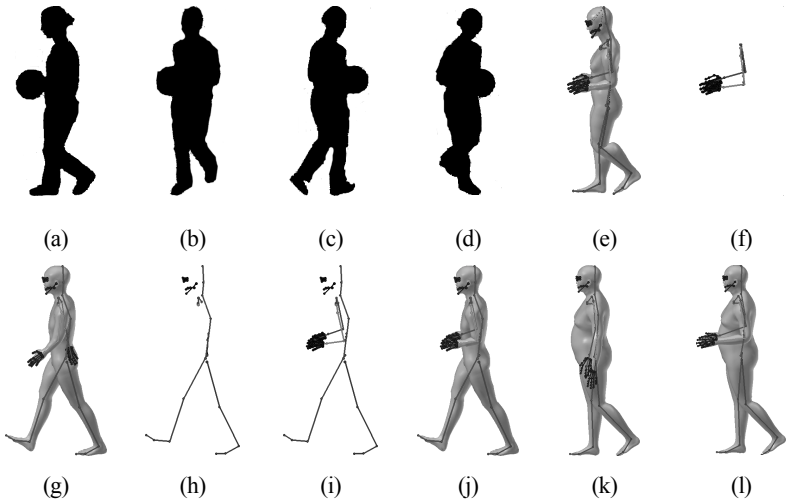


Figure 9. Synthesizing virtual gait model: (a)-(d): multiview gait data with a ball; (e) 3D reconstructed gait model with skeleton embedded; (f) *ROI* of carrying-item skeleton; (g) normal-walking 3D gait model with skeleton embedded; (h) ROI^{-1} of normal-walking skeleton; (i) synthesized skeleton using (f) and (h); (j) the carrying-item posture synthesized model using (i); (k) 3D gait model with fat body shape; and (l) the carrying-item posture synthesized model using the proposed method.

3.6 Complete 3D parametric template models

In order to conduct fast 3D parametric model reconstruction, a group of normal standard 3D gait models with different postures in a gait cycle are constructed first. Together with the extended carrying-item posture synthetic models, the 3D parametric template models are enlarged to enable fast 3D reconstruction under different conditions. The extended template models are denoted by $T_m^k = F_{3D}(S_T, \psi_{m,S}^k)$, where T labels the template parameters.

The multi-view gait data and the corresponding views are the priori knowledge for 3D reconstruction and multi-view gait training. However, in the recognition process, the gait views θ and the *mth* carrying conditions may be unknown. Thus, the parameters should be detected that correspond to the most similar 3D parametric template models from the template dataset. To estimate the gait view θ and the *mth* carrying condition, the GEIs at θ view are computed from the template 3D gait models, i.e., $\Upsilon_{m,\theta}^{GEI} = \frac{1}{K}\sum_{k=1}^{K} B_{m,\theta}^{k}$, where K is the number of gait template models within a gait period fixed by k-clustering, and $B_{m,\theta}^{k}$ denotes 2D silhouettes from the 3D template model at θ view and k phase of a gait period under *mth* carrying condition. The gait view directions are quantized into L directions. $\Upsilon_{m,\theta_l}^{GEI}$ is defined as the N_G dimension vector of GEIs gait feature of *mth* carrying condition re-projected onto view θ. The GEIs gait vectors with L view directions and M carrying conditions are then collected as training data set. Using the GEIs gait features associated with different carrying conditions, a matrix is constructed where the rows represent the view directions and the columns indicate the carrying conditions in the training data set which is decomposed by SVD, i.e.,

$$\Theta = \begin{bmatrix} \Upsilon^{GEI}_{1,\theta_1} \cdots \Upsilon^{GEI}_{M,\theta_1} \\ \mathrm{M\,O\,M} \\ \Upsilon^{GEI}_{1,\theta_L} \cdots \Upsilon^{GEI}_{M,\theta_L} \end{bmatrix} = USV^T = \begin{bmatrix} P_{\theta_1} \\ \mathrm{M} \\ P_{\theta_L} \end{bmatrix} \begin{bmatrix} v^1 \cdots v^M \end{bmatrix}, \tag{10}$$

where U is the $LN_G \times M$ orthogonal matrix, V is the $M \times M$ orthogonal matrix, S is the $M \times M$ diagonal matrix composed of singular values P_{θ_l}, $l = 1, 2, \ldots, L$, which is the $N_G \times M$ submatrix of US, v^m, $m = 1, 2, \ldots, M$, and M is the M dimensional column vector. The vector v is an intrinsic feature vector of the *mth* carrying condition that is independent of view directions. P_{θ_l} is the operator factor for carrying condition and is a projection matrix which is independent of the carrying condition. It projects the intrinsic vector v to the feature vector Υ for view direction θ, i.e.,

$$\Upsilon^{GEI}_{m,\theta_l} = P_{\theta_l} v^m. \tag{11}$$

Given a probe gait GEIs $\hat{\Upsilon}^{GEI}_{m,\theta_l}$ at θ_l view, the carrying condition identity coefficient is

$$\hat{v} = (P_{\theta_l})^{-1} \hat{\Upsilon}^{GEI}_{m,\theta_l}. \tag{12}$$

The maximum probability information of identity that is selected to achieve carrying condition is

$$p(p \mid \hat{v}) \infty \exp[-\left\| \hat{v} - v^m \right\|^2 / (2\sigma^2)]. \tag{13}$$

To automatically determine the *mth* carrying condition of the probe subject, the gait view θ_l must be estimated first. The SGEIs in [38] are introduced for arbitrary gait view estimation. SGEIs are obtained by cropping the region enclosed between the bottom of the GEIs and up to the anatomical positions of knee, i.e., 0.285H, where H is the height of GEI. SGEI is more effective for view detection than the method CCA [39] due to its ability to overcome the adverse effect of clothing and carried items based on anthropometric analysis. However, neither SGEIs based gait view detection method nor the method CCA is suitable for classifying the front and back views (i.e., at 0° and 180°, respectively) and views close to them (i.e., at 18° and 162°), as the shape characteristics of a subject remain almost similar in these cases. Also, the degradation in performance for views 72° and 108° is attributed to the similar shape characteristics with the subjects at 90° [7]. In order to overcome these problems, the silhouette size gradient score is introduced as an additional constraint for further classification of the estimated gait views. The silhouette size gradient score is defined as $Score = \sum_{n=1}^{N} \left| \nabla F_{size}(n) - G^{\theta}_{mean} \right|$, where $F_{size}(n)$ denotes the original gait silhouette size of *nth* frame in a gait period, N is total number of gait frames and G^{θ}_{mean} is mean silhouette size gradient of θ view (estimated using training data set). The original gait silhouette size is calculated from image pixels. The walking direction is indicated by $Dir = \sum_{n=1}^{N} I\left[\nabla F_{size}(n)\right]$ where $I[\cdot]$ is an indicator that is 1 if gradient is greater than 0 and -1 otherwise. If $Dir \geq 0$, the estimated gait view ranges from 0° to 90°, otherwise from 90° to 180°. This is due to the silhouette increasing in size while the subject walks towards the camera and the silhouette size gradient is greater than zero. When walking away from the camera, the situation is reversed. By our proposed method, the gait views can be first estimated using SGEIs and be further determined by $\arg\min_{\theta} Score$ and walking direction indicator. Using the gait silhouette size gradient and direction indicator,

0° and 180° that have the same shape characteristics are easily distinguished. 72° and 108° that are attributed to the similar shape characteristics differ significantly in the walking direction indicators.

3.7. Self-occlusion optimized simultaneous Sparse Representation Model

1) Gait sparse representation

The original goal of sparse approximation is to find a line combination of regressors to represent a signal vector for compression purpose. Wright et al. [40] exploited the discriminative nature of sparse representation for classification. They proposed a 2D face recognition method based on sparse representation from frontal views with varying expression and illumination, as well as occlusion and facial disguise. In sparse representation, a complete dictionary is necessary in which the elements are chosen to achieve sparse linear combination. The dictionary is usually constructed from standard bases (e.g., Fourier and Gabor atoms). However, in face, gait and other image-based recognition, the base elements of an over-complete dictionary are the training samples.

Given sufficient n_i training samples of the *ith* object class $A^i = [\gamma_1^i, \gamma_2^i, ..., \gamma_{n_i}^i] \in \Re^{m \times n_i}$, any probe sample $y \in \Re^m$ from the same class is approximated from the linear combination of the training samples associated with object *i* [40], i.e.,

$$y = x_1^i \gamma_1^i + x_2^i \gamma_2^i + ... + x_{n_i}^i \gamma_{n_i}^i = A_i x^i, \tag{14}$$

where $x_{n_i}^i \in \Re$ denotes the weighting coefficient and $x^i = [x_1^i, x_2^i, ..., x_{n_i}^i]$ is a vector composed by the weighting coefficient for each training sample belonging to *ith* object class. Since the class *i* is initially unknown for the probe object, the linear representation of *y* is redefined in terms of all training samples as

$$y = Ax = \left[A^1, A^2, ..., A^K\right]\left[0, ..., 0, x_1^i, x_2^i, ..., x_{n_i}^i, 0, ..., 0\right]^T, \tag{15}$$

where $A = [A^1, A^2, ..., A^K]$ represents the dictionary for the entire training set including the K sub dictionary A^K, and x denotes the sparse description vector whose entries are zero except those associated with the *ith* class. To realize the sparse representation of probe *y*, the sparse vector x is computed by solving

$$\hat{x} = \arg\min_x \| x \|_1 \quad \text{subject to} \quad \| Ax - y \|_2 \le \varepsilon, \tag{16}$$

where $\varepsilon > 0$ is an optional error tolerance. The probe *y* is classified as *ith* object class by minimizing the residual between *y* and $\hat{y}_i$, i.e.,

$$i = \arg\min_i \| y - A\delta_i(\hat{x}) \|_2 = \arg\min_i \| y - A\hat{x}^i \|_2, \tag{17}$$

where $\delta_i(x) \in \Re^n$ is a vector whose only nonzero entries are the entries in x that are associated with *ith* class.

2) Gait recognition based on Self-occlusion Optimized Simultaneous Sparse Representation model.

Fig. 10 shows the gait recognition process based on self-occlusion optimized simultaneous sparse representation model. First the over-complete dictionary A of 3D gait is constructed with variant-phase 3D gait models. Variant-phase gait models are reconstructed using the multi-views 2D gait silhouettes corresponding to different phase in a gait period as in Section 3.3. They are 3D parametric gait models with different postures.

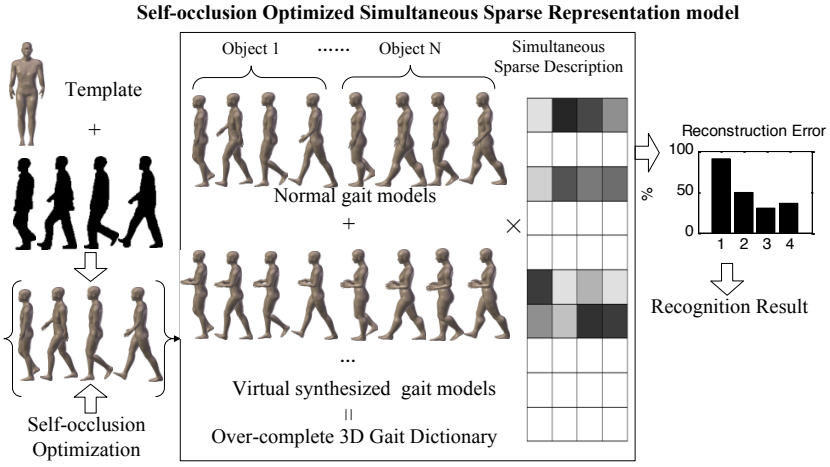


Figure 10. Gait recognition process based on simultaneous spare representation model.

According to the theory of sparse representation and compressed sensing, any probe model y can be represented by the linear combination of the atoms in dictionary if given sufficient training samples. In order to obtain sufficient training samples, the different virtual 3D gait models with variant-phase are synthesized in multiple carrying conditions. Together with the normal variant-phase 3D gait models, the over-complete dictionary is constructed.

For the parametric gait model, the unified 3D model is fast reconstructed by morphing using the estimated static shape features S and motion features ψ. The shape vector S and motion vector ψ are then combined to form the 3D gait vector $\gamma_{n_i}^i = \left\{ S_{n_i}^i, \psi_{n_i}^i \right\}$, where n_i denotes the total training samples of *ith* object with different postures. As a result, the *ith* object sub dictionary $A^i = [\gamma_1^i, \gamma_2^i, ..., \gamma_{n_i}^i]$ is constructed. The over-complete dictionary for the entire training set includes the K object class, where each class is composed of 3D normal gait models together with different carrying condition, and synthesized virtual gait models. It is denoted by

$$A = [A^1, A^2, ..., A^K] = [\gamma_1^i, \gamma_2^i, ..., \gamma_{n_i}^i, ..., \gamma_n^K] \in \Re^{m \times n}, \tag{18}$$

where $n = \sum_{i=1}^{K} n_i$ equals the total training samples. Given any 3D model y in a gait period, it can be sparsely represented by $y = Ax_0$, where x_0 denotes the sparse weighting coefficient vector.

Since a gait period is composed by several gait phases with different postures, it is sufficiently distinct to enable subject identification using a sufficient number of reconstructed 3D gait models in a period. The multi-phase gait models are then collected as input data set denoted by $Y = \{y_1, y_2, ..., y_M\}$, where M is the total number of input gait models in a gait period. The sparse representation problem is solved by

$$\{ \hat{x}_i \}_{i=1}^{M} = \arg\min_{\{x_i\}} \sum_{i=1}^{M} \| x_i \|_2 \tag{19}$$

subject to $\sum_{i=1}^{M} \| y_i - Ax_i \|_2 \le \varepsilon, \forall 1 \le i \le M$.

In 2D gait recognition, gait silhouettes in a period are usually averaged in various ways to form GEI [10], GFI [41], CGI [42], DNGR [43] or similar gait energy features for recognition. The averaged gait features are usually sensitive to the accurate detection of the gait period. If occlusion occurs resulting in only a few gait silhouettes or the subject deviates significantly from the expected straight trajectory, the recognition rate would decrease significantly. In order to overcome these problems, Self-Occlusion Optimized Simultaneous Sparse Representation Model

(SOO-SSRM) is introduced to achieve arbitrary view gait recognition that does not require a complete gait cycle data or assumes that the subject walks in a similar trajectory. The total number M of input gait models is not limited and the simultaneous sparse representation problem is solved by

$$\hat{X} = \arg\min_{X} \| X \| \tag{20}$$

subject to $\| YW - AX \|_F \le \varepsilon$. $X = [x_1, x_2, ..., x_M]$ and $\| \cdot \|_F$ denotes the Frobenius Norm. $W = [w_1, w_1, ..., w_M]^T$ is self-occlusion optimized matrix.

In the training process for AVGR-BRPS, the 3D body can be well reconstructed due to the use of multi-view gait images. For multiple input frames, the body shape is optimized and the skeleton structures are well estimated by overcoming the self-occlusion problems. However, in the recognition step, it is not practical in real applications to use numerous cameras or multi-view images to reconstruct the body. However, single view images provide only one-side surface portion of the human body. Using our clothes-independent 3D body and pose estimation method based on template gait data as prior knowledge, the gait models can be well reconstructed for gait recognition. In order to further improve the robustness of the algorithm, a self-occlusion optimized coefficient w_θ^i is introduced to optimize the estimated body shape and posture parameters. θ denotes the single probe gait view for recognition and it is associated with the partial occlusion of the body. Only the surface of θ is available and the other side is self-occlusion. The optimized body features are denoted by $y_i' = y_i w_\theta^i$. Using training multi-view gait data, the function of the coefficient w_i is estimated by

$$\arg\min_{w_i} \| \gamma_\Phi^i - \gamma_\theta^i w_\theta^i \|, \tag{21}$$

where γ_Φ^i is the 3D gait model vector estimated using multi-view gait images and Φ is a multi-view set. γ_θ^i denotes the 3D gait model vector using only a single θ view where $\theta \in \Phi$.

By estimating the sparse representation vector $\hat{X}$, the reconstructed residual is given by $\hat{i} = \arg\min_i r_i(Y) = \arg\min_i \left\| Y - A\delta_i(\hat{X}) \right\|_2$, where δ_i is the characteristic function that selects the coefficients associated with the *ith* class. The final classification of multi-input gait model Y is determined by $identity(Y) = \arg\min_i r_i(Y)$ with the minimized residual.

3) Optimized Strategy for Simultaneous Sparse Representation-based Classification

Since an over-complete dictionary of 3D gait database is constructed in the proposed workflow, when dealing with a large number of training gait database the search for the minimum reconstructed residual in sparse representation-based classification could be slow and requires large memory. In order to address these problems, the carrying condition of the input probe subject is first determined by the method in Section 3.6. The normal gait database or corresponding virtual synthesized gait database is selected for the next matching process. Second, in order to reduce the large samples, the 3D body shape vector S is used to fast index K *(K<M)* candidate objects in the selected dictionary that are similar by Euclidean distance. The efficient feature-sign search algorithm [44] is then used to speed up the sparse coding, which enables larger sparse codes to be learned. Given a group of input probe 3D models in a gait period that are denoted by $y^i = \{\gamma_{n_i}^i\} = \left\{(S_{n_i}^i, \psi_{n_i}^i)\right\}$,

their sparse coefficients $x_{n_i}^i$ can be exploited in dataset that only includes K candidate models.

4 Experiments

In our proposed gait recognition method, accurate 3D parametric gait model needs to be constructed from multi-view 2D gait silhouettes. Table 1 shows the most popular multi-view gait datasets (with 3 viewpoints at least) for evaluating gait recognition methods. Multi-view gait data captured by multi-cameras at similar time with multiple covariate factors are also necessary. These two requirements meant the two SOTON databases [53,54], OU-ISIR Treadmill C [55], and WOSG [5] are not appropriate. Also the WOSG dataset was generated using one camera, which is not suitable for 3D model reconstruction. CASIA B [3] comprises large view variations with 18-degree interval that is also under varied challenging conditions such as varied clothing (normal or coat), different carried items (backpack or handbag), etc. CMU MoBo database [49] comprises variations in speed, carrying status and tilt views. In CASIA B data acquisition, 11 cameras around the left hand side of the subject were used. In MoBo data capture, six cameras evenly distributed around the subject on the treadmill generated sufficient data for a full 3D reconstruction. These two databases are thus suitable for evaluating the proposed method.

Table 1 Popular multi-view gait datasets involving at least 3 views.

Gait Database	Time	Subjects	Covariates	Views
CMU MoBo [49]	2001	25	Viewpoint, Speed, Carrying items	6
SOTON Multimodal Database [53]	2011	>300	Viewpoint	12
SOTON Temporal Database [54]	2012	25	Viewpoint, Time difference	12
CASIA B [3]	2006	124	Viewpoint, Clothing, Carrying items	11
OU-ISIR Treadmill C [55]	2010	200	Viewpoint	25
WOSG [5]	2013	155	Viewpoint, Illumination	8

4.1. Experiments on CASIA B dataset

CASIA Dataset B is a multiview gait dataset comprising 124 subjects, and the gait data was captured from 11 views (as illustrated in Fig. 11) in the range [0° 180°] with an interval of 18°. 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 [3]. The video sequences are recorded indoor at a rate of 25 frames per second and the resolution of each frame is 320×240.

Figure 11. CASIA DataSet illustrations.

In our experiments, the first four normal-walking sequences with 11 views of all subjects are considered as gallery. The first set of walking sequences with carrying conditions of 1-10 subjects is considered as carrying items training gallery. The remaining two normal-walking sequences, two walking sequences wearing a coat and walking sequences with three variations carrying conditions are considered as the probe.

1) Estimation of gait views with different methods.

Fig. 12 – Fig. 14 show the comparisons of correct view matching rate (CVR) on CASIA B dataset for subjects walking normally, walking with a bag and walking wearing a coat. Our method based on SGEIs with silhouette size gradient score achieves the highest CVR for views 0° and 180°, and views close to them (i.e., 18° and 162°). The CCA [39] method which uses Gaussian process is not suitable for estimating the front and back views (i.e., 0° and 180°) and has low CVR for 90°. The low CVR for views 0° and 180°, as the shape characteristics of a subject remain almost the same in these cases, has been greatly improved compared with VI-MGR proposed in [7]. The silhouette size gradient score as view range constraints for final classification of the estimated gait views is helpful in improving the CVR. In support of this observation, a confusion matrix for CVR of our method corresponding to normal walking cases of probe subjects is presented in Table 2. Similar to VI-MGR, only 22 subjects' SGEIs together with their silhouette size gradient scores are used for correct view matching while CCA uses 60% of the total subjects as training. The SGEIs based method is more robust than CCA method.

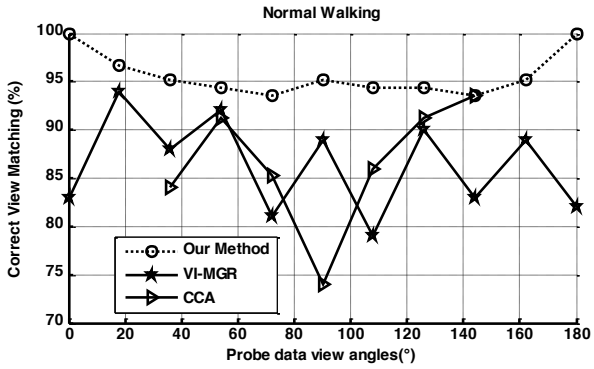


Figure 12. Correct view matching of normal walking.

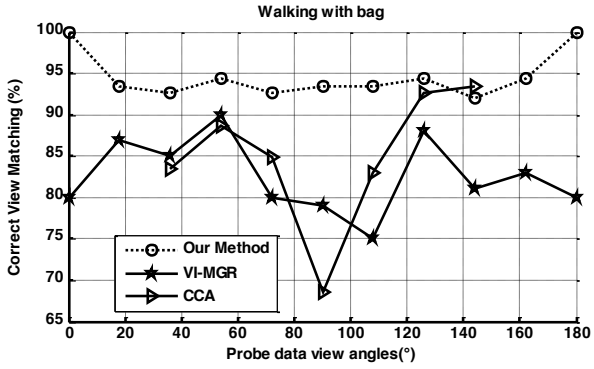


Figure 13. Correct view matching of walking with bag.

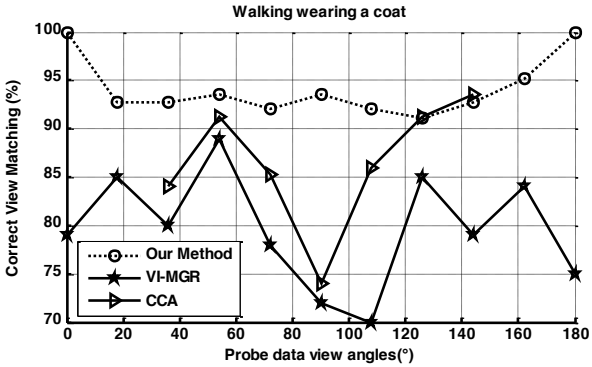


Figure 14. Correct view matching of walking wearing coat.

Table 2 Confusion matrix for CVR of our method on the normal walking subjects of CASIA B gait dataset. VI-MGR data are enclosed in parentheses.

View	Estimated view										
	0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°
0°	124(102)	0	0	0	0	0	0	0	0	0	0(22)
18°	0	120(116)	4(6)	0	0	0	0	0	0(1)	0(1)	0
36°	0	0	118(109)	6(10)	0(3)	0	0	0(2)	0	0	0
54°	0	0	2	117（114）	5(7)	0(1)	0	0	0	0	0
72°	0	0	0	4	116(100)	4(5)	0(5)	0	0	0	0
90°	0	0	0	0	3(8)	118（110）	3(6)	0	0	0	0
108°	0	0	0	0	0(6)	4(6)	117(98)	3(4)	0	0	0
126°	0	0	0	0	0	0	5(10)	117（112）	2	0	0
144°	0	0	0	0	0	0	0	3(9)	116（103）	5(12)	0
162°	0	0	0(4)	0	0	0	0	0(2)	6(8)	118(110)	0
180°	0(22)	0	0	0	0	0	0	0	0	0	124(102)

2) Multi-view gait recognition in various conditions

In order to evaluate the robustness of our AVGR-BPRS in various conditions under multi-views, we compare our method with VI-MGR [7], CCA [39], RLTDA [45], GEI-SVD [46], Robust VTM [47] using CASIA Dataset B. Fig. 15 and Fig. 16 show the rank-1 recognition rates of AVGR-BPRS and VI-MGR for walking subjects with a bag and wearing a coat in views from 0° to 180°. For CCA method the view angles are 36°, 54°, 90°, 108°, 126° and 144° for subjects carrying a bag and wearing a coat using GFI as a gait signature.

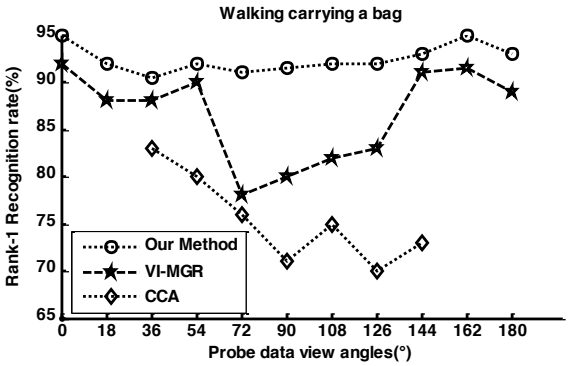


Figure 15. Rank-1 recognition rates of AVGR-BPRS, VI-MGR and CCA on walking subjects with a bag.

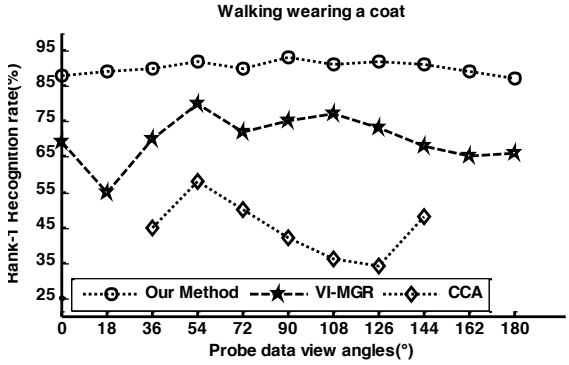


Figure 16. Rank-1 recognition rates of AVGR-BPRS, VI-MGR and CCA on walking subjects wearing a coat.

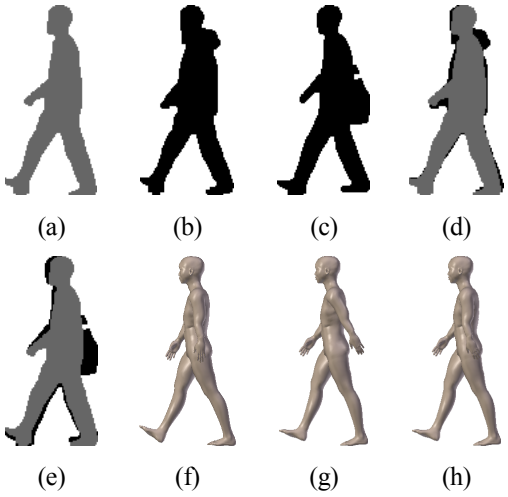
Table 3 shows the rank-1 recognition rate of AVGR-BPRS, method [30], RLTDA, GEI-SVD and Robust VTM on CASIA B gait dataset with 54°, 90°and 126° views. Among these methods, the known probe gait feature from one view angle is transformed into the most similarity gallery view using a VTM for similarity matching. We removed all the 2D silhouettes at probe view in the 3D reconstructed process. This is because, in VTM based methods, the chosen probe viewing angle must be excluded from the other gallery viewing angles that are used for training in order to evaluate the robustness of view transform model. The multi-view gait silhouettes not at probe view are used for training.

Table 3 Rank-1 recognition rates of AVGR-BPRS, RLTDA, GEI-SVD and Robust VTM on CASIA B gait dataset.

Probe	Our Method		Method [30]		RLTDA		GEI-SVD		Robust VTM	
	Bag	Coat	Bag	Coat	Bag	Coat	Bag	Coat	Bag	Coat
54°	90	92	76.4	87.9	80.8	69.4	31.4	17.7	40.7	35.4
90°	92	93	73.7	91.1	76.5	72.1	32.1	29.0	58.2	50.3
126°	91	92	76.9	86.2	72.3	64.6	44.1	34.6	59.4	61.3

AVGR-BPRS outperforms all the existing methods and the most obvious reason is illustrated in Fig. 17, which shows that AVGR-BPRS 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 constraint. The bag not belonging to the body is removed in the 3D morphing process. However, in some cases, the bag could influence the body reconstruction. Fig. 17(g) shows that without using carrying-item posture synthesized model the reconstruction of the left hand is affected by the bag silhouette. Fig. 17(h) shows the correct 3D morphing result using carrying-item posture synthesized model.

Fig. 15 shows that the existing methods achieve high gait recognition rates when the views are similar to 0° and 180°, and achieve low rates when the views are closer to 90°. This is because the gait silhouettes with a bag at 90° provide the most information of the carrying conditions. By using combined multi-view gait silhouettes for 3D parametric model reconstruction in AVGR-PBRS significantly reduces the influence of the carrying conditions compared with situations just using disturbed gait silhouettes for recognition.



607 Figure 17. 3D body reconstruction: (a) normal walk; (b) walk wearing a coat; (c) walk with a bag; (d) contour difference between
608 (a) and (b); (e) contour difference between (a) and (c); (f) reconstruction from (b); (g) reconstruction from (c) without using
609 carrying-item posture synthesized model; and (h) reconstruction from (c) using carrying-item posture synthesized model.

610

611 Table 3 shows that the existing methods based on VTM or view-independent models have low recognition rates
612 in various carrying conditions. It indicates that the VTM or view-independent models trained by normal gait dataset
613 are not robust to clothing and carrying conditions when the viewing angle is changed.

614 Additional experiments were conducted to show the advantages of our method. Table 4 illustrates the rank-1
615 recognition rate of AVGR-BPRS on CASIA B gait dataset with different size of training views. Note that
616 AVGR-BPRS can realize multi-view gait recognition just using one view gait dataset for training. However a larger
617 number of training samples with various carrying conditions and view angles will contribute to a more accurate
618 reconstructed 3D parametric gait models and an over-complete dictionary of 3D gaits, leading to even better
619 performance in arbitrary view gait recognition under various conditions. Table 4 shows that where the probe views
620 are included in the training gallery views, the recognition rate is higher. This is why better recognition rates are
621 obtained for 90 and 162 degree in condition C than in B.

622 Table 4 Rank-1 recognition rate of the AVGR-BPRS on CASIA B gait dataset with different views for training. A: training with
623 11 views from 0° to 180°; B: training with 6 views including 0°, 36°, 72°, 108°, 144° and 180°; and C: training with 3
624 views including 18° ,90° and 162°.

View	Recognition rate (%)								
	Normal walking			Walking with a bag			Walking wearing a coat		
	A	B	C	A	B	C	A	B	C
18°	99	93	92	92	90	90	89	84	84
36°	100	98	90	91	90	88	90	88	81
54°	98	92	86	90	89	85	92	85	80
72°	99	95	91	91	90	86	90	87	82
90°	100	94	95	92	88	91	93	86	90
108°	100	96	92	92	91	88	91	88	81
126°	98	93	88	91	90	84	92	85	80
144°	99	97	90	93	92	86	91	87	83
162°	98	92	93	95	90	90	89	85	86

625

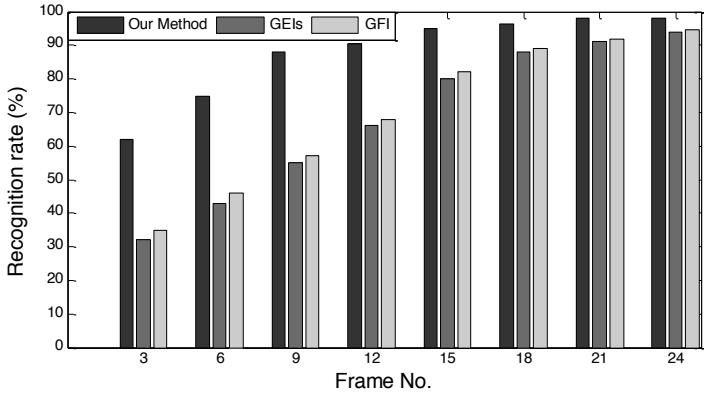


Figure 18. Rank-1 recognition rates of AVGR-BPRS, GEIs and GFI for walking normally in 90° view with random number of probe frames.

3) Gait recognition with random number of probe frames

Fig. 18 shows the rank-1 recognition rates of AVGR-BPRS, GEIs [10] and GFI [41] for walking normally at 90° view with random number of probe frames. Most existing methods of gait recognition assume that the gait cycle is well estimated and the complete gait cycle frames are collected. However, the gait cycle is sometimes not so easy to obtain especially in occlusion conditions or key pose gait frames are missing. In those cases, the limited gait frames are used for recognition. Most existing methods especially those based on averaged silhouettes as features (e.g., GEI, GFI, PEI [48]) have unsatisfactory performances due to the lack of a complete cycle gait silhouettes. Using simultaneous sparse representation model enables AVGR to achieve the high recognition rate compared with existing methods. Unlike the statistical approaches, the 3D model-based gait recognition is robust when facing such conditions.

4) Computational Complexity

Sparse representation-based classification stage is the time-consuming part of gait recognition with virtual synthesized over-complete gait dataset. Thus, we discuss the computational complexity of the spare representation coding and the minimum reconstructed residual search involved in the proposed method. The time complexity of original method is $O(MInV)$ if 3D down-sampled vertices are used directly as features prior to the optimized strategy. M is the number of the virtual synthesized dataset with different carried item posture, I denotes the number of total class, n is the total training samples in a gait period with different carrying-item postures and V is the vertices number of 3D gait mesh model. The time complexity of our optimized method is $O(KnL)$, where K $(K<M)$ indicates the candidate objects selected from total M objects in dictionary, L defines the length of the 3D gait encoding feature $\gamma = \{S, \psi\}$, which is constructed by discretised semantic values of body shape S and posture parametric of BVH. Table 5 shows the typical running time of sparse representation based classification stage on a PC with an Intel Core 2.93GHz CPU and 2GB RAM. In Table 5, K is 20% size of I and M *is* set to three with one normal gait dataset and two virtual synthesized gait datasets, i.e., bag carrying and ball carrying. The running time in our sparse representation based classification might be slower than the 2D based multi-view gait recognition. This is because unlike the GEI or similar method that average silhouettes across a gait cycle to represent gait features, our method reconstructs each 3D model across a gait cycle and search the minimum reconstructed sparse representation residual for classification in an over-complete dictionary with virtual synthesized gait

datasets. However, our method is more robust to achieve higher view-invariant recognition with random number of probe frames and the subjects do not need to walk in a similar trajectory. It shows the potential in real-world applications.

Table 5 Typical running time of sparse representation-based classification stage.

Classification methods	Complexity	Average time(seconds)
Optimized strategy method	$O(KnL)$	8.2
Original method	$O(MInV)$	122.6

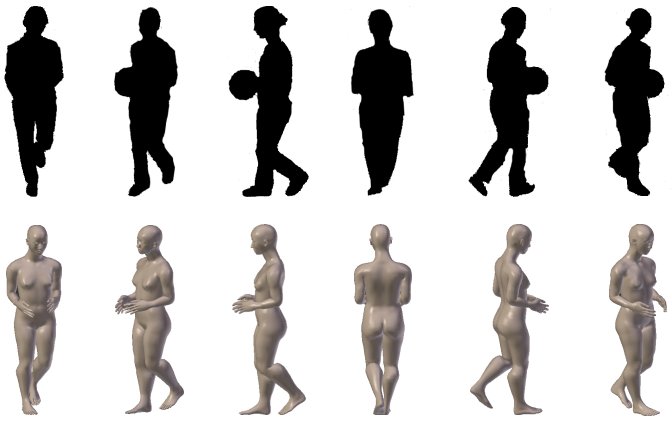
4.2. Experiments on CMU Motion of Body (MoBo) Database

The CMU MoBo database [49] consists of six sequences of 25 subjects (23 males, 2 females) walking on a treadmill. The 3 CCD progressive scan images have a resolution of 640×480. Each subject is recorded performing four different types of walking: slow walk, fast walk, inclined walk, and slow walk holding a ball (to inhibit arm swing). Each sequence is 11 seconds long and recorded at 30 frames per second.

Nine experiments were performed with steps as shown in Table 6 for robust test. Since the dataset is not too large, 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) [48]. In order to evaluate our proposed method for robustness, experiments are chosen with gallery and probe sets under various conditions. Since our virtual model synthesized model needs a template for carrying BVH data extraction, one sequence with ball as illustrated in Fig. 19 is chosen as the gallery and other walk sequences with ball as the probe set.

Table 6 Nine experiments on CMU MoBo gait dataset (in lateral view).

Experiment	Gallery set	Probe set
S vs. F	Slow walk	Fast walk
S vs. B	Slow walk	Ball walk
S vs. I	Slow walk	inclined walk
F vs. S	Fast walk	Slow walk
F vs. B	Fast walk	Ball walk
F vs. I	Fast walk	Inclined walk
I vs. S	Inclined walk	Slow walk
I vs. F	Inclined walk	Fast walk
I vs. B	Inclined walk	Ball walk



(a) 0° (b) 45° (c) 90° (d) 180° (e)225° (f) 315°

Figure 19. Top row: gait silhoutttes and bottom row: Results of carried item models.

675 The approaches of STM-SPP [50], WBP [51], SGRVDL [52], Mehtod [19] and PEI [48] evaluated under varied
676 challenging conditions using CMU MoBo database are chosen for performance comparison. Table 7 shows that our
677 AVGR-BPRS outperforms the other methods especially for ball carrying condition and inclined walk. Results of the
678 experiments that are not presented in the original papers have been left blank in the table. The existing methods show
679 high recognition results when the gait silhouettes have small shape changes (e.g., S vs. F, F vs. S scenarios). However,
680 most approaches are not robust enough to appearance changes (e.g., F vs. B, I vs. B scenarios) due to the shortage of
681 2D gait silhouette based methods. The 2D gait silhouettes could be influenced easily by variant carrying conditions. In
682 contrast, the performance of our 3D gait model based algorithm shows satisfactory classification results across all
683 types of gallery/probe conditions.

684 There are several reasons why our AVGR-BPRS achieves significantly better performance. The first is that our
685 reconstructed 3D models are based on unified parametric body model and multi-view 2D gait silhouettes. It makes
686 our method more efficient dealing with appearance changes and wearing conditions. The second is the use of
687 carrying-items posture synthesized model. The virtual 3D gait models that are not in the training dataset are
688 synthesized using prior BVH data. The last is that 3D parametric model is less sensitive to inclined silhouettes. This is
689 because the 3D models can be manipulated around X-Y-Z axes to fit any changes in hierarchy and initial pose of the
690 body whereas the 2D silhouettes cannot. As the hierarchy and initial pose of the skeleton are included in the motion
691 features ψ , the inclined conditions are considered in 3D pose estimated steps. The above advantages make our
692 proposed method especially suitable for surveillance applications where various wearing and carrying conditions
693 influence the appearance based gait silhouettes.

694 Table 7 Recognition results on Mobo data set using different methods.

Experiment	STM-SPP	WBP	SGRVDL	Method [19]	PEI	Our Method
S vs. F	94%	92%	96%	92%	100%	96%
S vs. B	93%	73%	87%	-	92%	94%
S vs. I	-	-	-	-	60%	88%
F vs. S	91%	92%	92%	92%	88%	92%
F vs. B	84%	61%	88%	-	60%	93%
F vs. I	-	-	-	-	72%	88%
I vs. S	-	-	-	-	76%	91%
I vs. F	-	-	-	-	80%	88%
I vs. B	-	-	-	-	48%	86%
B vs. S	82%	75%	87%	-	92%	92%
B vs. F	82%	63%	88%	-	84%	91%
B vs. I	-	-	-	-	76%	86%

695 **5 Conclusions**

696 In this paper, a novel 3D model-based arbitrary view gait recognition method, AVGR-BPRS, is proposed.
697 AVGR-BPRS is robust to variations in speed, inclined plane, clothing and presence of a carried item, which might be
698 encountered in real scenarios. The experimental results show that AVGR-BPRS is more effective than existing
699 multi-view gait recognition approaches in the case of multiple covariates (e.g., with a coat, carrying a bag, walk with
700 a ball, and walk inclined).

701 AVGR-BPRS that is based on clothes-independent 3D statistical human body model generated from multiview
702 2D gait silhouettes achieves high performance rate with limited training gait views. It achieves arbitrary view gait
703 recognition without any camera calibration. Using the proposed carrying-item posture synthesized model, different

virtual gait models are obtained by embedding the extracted carrying-item skeleton BVH data into the standard training 3D gait models. This greatly affects the 3D gait itself rather than mitigating the effect of changes in covariate conditions that affect gait feature extraction in existing 2D methods. The self-occlusion optimized simultaneous sparse representation model provides a new approach for achieving arbitrary view gait recognition that does not require a complete gait cycle data or assumes that the subject walks in a similar trajectory.

Acknowledgments

This work was supported partly by the National Natural Science Foundation of China (Grant No.61403426, No.91220301), the Scientific Research Project of Hunan Provincial Education Department, China (Grant No.15C0981), the Science and Technology Key Program of Hunan, China (Grant No. 2015WK3006).

References

[1] I. Bouchrika, M. Goffredo, J. Carter, M. Nixon. On Using Gait in Forensic Biometrics. Journal Of Forensic Sciences. 56(2011)883-889.

[2] S. Sarkar, P.J. Philips, Z. Liu, I. Vega, P. Grother, K. Bowyer, The HumanID gait challenge problem: data sets, performance, and analysis, IEEE Trans. Pattern Anal. Mach. Intell. 27 (2) (2005) 162–177.

[3] S. Yu, D. Tan, T. Tan, A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition, in: Proceedings of the 18th International Conference on Pattern Recognition, 2006, pp 441–444.

[4] J.D. Shutler, M.G. Grant, M.S. Nixon, and J.N. Carter, On a Large Sequence-Based Human Gait Database. Proc. 4th International Conf. on Recent Advances in Soft Computing, Nottingham (UK), 2002:pp66-71.

[5] B. DeCann, A. Ross, J. Dawson. Investigating Gait Recognition in the Short-Wave Infrared (SWIR) Spectrum: Dataset and Challenges. Proceedings of the SPIE - The International Society for Optical Engineering, 2013, pp1-16.

[6] H. Iwama, M. Okumura, Y. Makihara, Y. Yagi, The OU-ISIR Gait Database Comprising the Large Population Dataset and Performance Evaluation of Gait Recognition, IEEE Trans. on Information Forensics and Security, 2012(7)1511-1521.

[7] S.D. Choudhury, T. Tjahjadi. Robust view-invariant multiscale gait recognition. Pattern Recognition. 48 (2015) 798–811.

[8] J. Gu, X. Ding, S. Wang, Y. Wu. Action and Gait Recognition from Recovered 3-D Human Joints. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics. 40(4) (2010) 1021-1033.

[9] S. Sivapalan, D. Chen, S. Denman, S. Sridharan, C. Fookes. Gait energy volumes and frontal gait recognition using depth images, in Proc. of the 2011 International Joint Conference on Biometrics, Piscataway, 2011:pp1-6.

[10] J. Han, B. Bhanu, Individual recognition using gait energy image, IEEE Trans. Pattern Anal. Mach. Intell. 28 (2) (2006) 316–322.

[11] G. Zhao, G. Liu, H. Li. M. Pietikainen. 3D gait recognition using multiple cameras. In Proc. of the 7th International Conf. on Automatic Face and Gesture Recognition, Southampton, United Kingdom, April 2006, pp.529-534.

[12] G. Rogez, J. Rihan, J.J. Guerrero, C. Orrite. Monocular 3-D Gait Tracking in Surveillance Scenes. IEEE Transactions on Cybernetics.44 (2014) 894-909.

[13] G. Shakhnarovich, L. Lee, T. Darrell. Integrated face and gait recognition from multiple views. In Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, Kauai, HI, USA, 8–14 December 2001; pp. 439–446.

[14] J. Tang, J. Luo, T. Tjahjadi, Y. Gao. 2.5D Multi-View Gait Recognition Based on Point Cloud Registration. Sensors, 14 (2014) 6124-6143.

[15] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, Y. Yagi. Gait recognition using a view transformation model in the frequency domain. In Proc. of the 9th European Conf. on Computer Vision, Graz, Austria, 7–13 May 2006, pp. 151–163.

[16] W. Kusakunniran, Q. Wu, J. Zhang, H. Li. Gait recognition under various viewing angles based on correlated motion regression. IEEE Trans. Circuits Syst. Video Technol, 22 (2012) 966–980.

[17] W. Kusakunniran, Q. Wu, J. Zhang, H. Li. Cross-view and multi-view gait recognitions based on view transformation model using multi-layer perceptron. Pattern Recognit. Lett. 33 (2012) 882–889.

[18] D. Muramatsu, A. Shiraishi, Y. Makihara, M.Z. Uddin, and Y. Yagi. Gait-Based Person Recognition Using Arbitrary View Transformation Model. IEEE Transactions on Image Processing. 24(1) (2015) 140-154.

[19] Z. Wei, W. Cong.View-invariant gait recognition via deterministic learning. Neurocomputing, 175(2016)324-335.

[20] F. Jean, A.B. Albu, R. Bergevin. Towards view-invariant gait modeling: Computing view-normalized body part trajectories. Pattern Recognition. 42(2009) 2936–2949.

[21] M. Goffredo, I. Bouchrika, J.N. Carter, et al. Self-calibrating view-invariant gait biometrics. IEEE Transactions on Systems Man & Cybernetics Part B Cybernetics A Publication of the IEEE Systems Man & Cybernetics Society, 40(4) (2010) 997 - 1008.

[22] W. Kusakunniran, Q. Wu, J. Zhang, H. Li. A New View-Invariant Feature for Cross-View Gait Recognition. IEEE Transactions on Information Forensics & Security, 8(10) (2013) 1642-1653.

[23] M. Hu, Y. Wang, Z. Zhang, J.J. Little, D. Huang. View-Invariant Discriminative Projection for Multi-View Gait-Based Human Identification. IEEE Transactions on Information Forensics and Security. 8 (2013) 2034-2045.

[24] H. Hu. Multiview Gait Recognition Based on Patch Distribution Features and Uncorrelated Multilinear Sparse Local Discriminant Canonical Correlation Analysis. IEEE Transactions on Circuits and Systems for Video Technology. 24 (2014) 617-630.

[25] C. Chen, J. Liang, H. Zhao, H. Hu, J. Tian. Frame difference energy image for gait recognition with incomplete silhouettes. Pattern Recognition Letters. 30(11) (2009) 977-84.

[26] M.A. Hossain, Y. Makihara, J. Wang, Y. Yagi. Clothing-invariant gait identification using part-based clothing categorization and adaptive weight control. Pattern Recognition, 43(6) (2010) 2281-91.

[27] Z. Xue, D. Ming, W. Song, B. Wan, S. Jin. Infrared gait recognition based on wavelet transform and support vector machine. Pattern Recognition, 43 (8) (2010) 2904-2910.

[28] K. Bashir, T. Xiang, S. Gong. Gait recognition without subject cooperation. Pattern Recognition Letters. 31 (13) (2010) 2052-2060.

[29] P. Chattopadhyay, S. Sural, J. Mukherjee, Frontal Gait Recognition From Incomplete Sequences Using RGB-D Camera. IEEE Transactions on Information Forensics and Security. 9(11) (2014) 1843-56.

[30] I. Rida, X. Jiang, G.L. Marcialis. Human Body Part Selection by Group Lasso of Motion for Model-Free Gait Recognition. IEEE Signal Processing Letters, 23(2016)154-158.

[31] C. Xin, X. Jiaming. Uncooperative gait recognition: Re-ranking based on sparse coding and multi-view hypergraph learning. Pattern Recognition, 53(2016)116-129.

[32] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, J. Davis. SCAPE: Shape Completion and Animation of People. ACM Transactions on Graphics. 24 (3) (2005) 408-416.

[33] B. Allen, B. Curless, Z. Popovic. The space of human body shapes: reconstruction and parameterization from range scans. ACM Transactions on Graphics. 22 (3) (2003) 587-94.

[34] N. Hasler, C. Stoll, M. Sunkel, B. Rosenhahn,, H.-P. Seidel, A Statistical Model of Human Pose and Body Shape. Computer Graphics Forum, 28 (2) (2009) 337-346.

[35] A. O. Bălan, M. J. Black The naked truth: Estimating body shape under clothing. Lecture Notes in Computer Science, 16(12) (2008) 1-14.

[36] N. Hasler, C. Stoll, B. Rosenhahn, T. Thormählen, H. P. Seidel. Estimating body shape of dressed humans. Computers & Graphics, 33(3) (2009) 211-216.

[37] S. Wuhrer, L. Pishchulin, A. Brunton, C. Shu, J. Lang. Estimation of human body shape and posture under clothing. Computer Vision & Image Understanding, 127(10) (2014) 31–42.

[38] S.D. Choudhury, T. Tjahjadi. Gait recognition based on shape and motion analysis of silhouette contours. Computer Vision and Image Understanding. 117 (12) (2013) 1770-1785.

[39] K. Bashir, T. Xiang, S. Gong, Cross-view gait recognition using correlation strength, in: Proceedings of the 21st British Machine Vision Conference, United Kingdom, 2010, pp. 109.1–109.11.

[40] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Y. Ma. robust face recognition via sparse representation. IEEE Transactions on Pattern Analysis and Machine Intelligence. 31(2) (2009) 210-227.

[41] T.H.W. Lam, K.H. Cheung, J.N.K. Liu. Gait flow image: a silhouette-based gait representation for human identification. Pattern Recognition. 44(4) (2011) 973-987.

[42] C. Wang, J. Zhang, L. Wang, J. Pu, X. Yuan, Human identification using temporal information preserving gait template, IEEE Trans. Pattern Anal. Mach. Intell. 34(11) (2012) 2164–2176.

[43] Z. Liu, S. Sarkar, Improved gait recognition by gait dynamics normalization, IEEE Trans. Pattern Anal. Mach. Intell. 28 (6) (2006) 863–876.

[44] L. Honglak, B. Alexis, R. Rajat, Y.Ng Andrew. Efficient sparse coding algorithms. Advances in Neural Information Processing Systems, 19(2006):801-808.

[45] Hu. Enhanced gabor feature based classification using a regularized locally tensor discriminant model for multiview gait recognition.IEEE Transactions on Circuits and Systems for Video Technology. 23 (7) (2013) 1274-1286.

[46] W. Kusakunniran, Q. Wu, H. Li, and J. Zhang. Multiple view gait recognition using view transformation model based on optimized gait energy images. in Proc. 12th ICCV Workshops, 2009, pp. 1058–1064.

[47] S. Zheng, J. Zhang, K. Huang, R. He, and T. Tan, Robust view transformation model for gait recognition, in Proc. Int. Conf. Image Process., 2011, pp. 2073–2076

[48] A. Roy,S. Sural,J. Mukherjee. Gait recognition using Pose Kinematics and Pose Energy Image .Signal Processing. 92 (3) (2012) 780-92.

[49] R. Gross, J. Shi, The CMU Motion of Body (MoBo) Database, Technical Report CMU-RI-TR-01-18, Robotics Institute, Carnegie Mellon University, 2001.

[50] S.D. Choudhury, T. Tjahjadi, Silhouette-based gait recognition using Procrustes shape analysis and elliptic Fourier descriptors, Pattern Recognition. 45 (2012)3414–3426.

[51] W. Kusakunniran, Q. Wu, H. Li, J. Zhang, Automatic gait recognition using weighted binary pattern on video, in: Proceedings of the Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance, 2009, pp. 49–54.

[52] W. Zeng, C. Wang. Silhouette-based gait recognition via deterministic learning. Advances in Brain Inspired Cognitive Systems. 6th International Conference, BICS 2013. Proceedings: LNCS 7888, 2013, pp.1-10.

[53] R.D. Seely, S. Samangooei, L. Middleton, J.N. Carter, M.S. Nixon. The University of Southampton Multi-Biometric Tunnel and Introducing a Novel 3D Gait Dataset. The 2nd IEEE International Conference on Biometrics: Theory, Application and System. Washington, USA. 2008, pp.1-6.

[54] D.S. Matovski, M.S. Nixon, S. Mahmoodi, J.N. Carter. The effect of time on gait recognition performance. IEEE Transactions on Information Forensics & Security, 7(2) (2012) 543-552.

[55] Y. Makihara, H. Mannami H, Y. Yagi. Gait Analysis of Gender and Age Using a Large-Scale Multi-view Gait Database, in 10th Asian Conference on Computer Vision, Queenstown, New Zealand, November 8-12, 2010, pp.975-986.