Recognition of Walking Humans in 3D: Initial Results

Koichiro Yamauchi[1,2] Bir Bhanu[1] Hideo Saito[2]
[1]Center for Research in Intelligent Systems
University of California, Riverside, CA 92521, USA
bhanu@cris.ucr.edu
[2]Keio University, Yokohama, 223-8522, JAPAN
{yamauchi, saito}@hvrl.ics.keio.ac.jp

Abstract

It has been challenging to recognize walking humans at arbitrary poses from a single or small number of video cameras. Attempts have been made mostly using a 2D image/silhouette-based representation and a limited use of 3D kinematic model-based approaches. In this paper, the problem of recognizing walking humans at arbitrary poses is addressed. Unlike all the previous work in computer vision and pattern recognition the models of walking humans are built using the sensed 3D range data at selected poses without any markers. An instance of a walking individual at a different pose is recognized using the 3D range data at that pose. Both modeling and recognition of an individual are done using the dense 3D range data. The proposed approach first measures 3D human body data that consists of the representative poses during a gait cycle. Next, a 3D human body model is fitted to the body data using an approach that overcomes the inherent gaps in the data and estimates the body pose with high accuracy. A gait sequence is synthesized by interpolation of joint positions and their movements from the fitted body models. Both dynamic and static gait features are obtained which are used to define a similarity measure for an individual recognition in the database. The experimental results show high recognition rates using our range based 3D gait database.

1. Introduction

Biometric authentication based on human gait has attracted significant attention recently. Gait has unique characteristics and indicates a personal trait. If we change the habit of walking consciously, the motion of body parts seem unnatural. Thus, gait involves not only the surface geometry but also dynamic motions of the joints.

Single-camera based gait recognition methods have been developed. Usually, they use a silhouette image extracted by background subtraction. Liu and Sarkar [1] generated hidden Markov model-based gait representation by estimating the stance state and then averaging silhouette images. Goffredo *et al.* [2] introduced view-independent markerless gait analysis based on the anthropometric propositions of human limbs and the characteristics of gait. To overcome the non-frontal pose problem more recently multi-camera based gait recognition methods have also been developed. These methods exhibited higher recognition accuracy for multi views than that of a single view. Zhao *et al.* [3] proposed gait tracking and recognition by matching 3D body model to video sequences. Gait feature is defined by the lengths of key segments and the motion trajectory of lower limbs which is obtained from the inferred model. Huang *et al.* [4] investigated the exploitation of the availability of multi views in a gait recognition system. The combination of the results of different views is evaluated to find the improvement in the recognition accuracy. Seely *et al.* [5] developed the University of Southampton Multi-Biometrics Tunnel. The subject's gait is recorded by eight synchronized cameras, and also the face and ear are captured by two videos separately.

Typically, gait recognition methods employ 3D pose inference, and then extract a gait feature. The effectiveness and correctness of the inference and extraction need to be verified before recognition. To overcome this problem, full body motion capture methods based on multi-camera system directly fit 3D body model to 3D body data. Cheung *et al.* [6] proposed temporal Shape-From-Silhouette algorithms for human body modeling and motion tracking. The markerless tracking algorithm is based on the Visual Hull alignment algorithm. Caillette *et al.* [7] proposed variable length Markov model which formulates human activities. Appearance model represented by Gaussian blobs is fitted onto voxels data from multiple views. The multi-camera system, 3D Room [8], acquires ten thousands triangles in the mesh to represent a whole body. When another object is measured, the size of voxels used is 7.8 mm$\times$7.8 mm$\times$7.8 mm. If we apply the multi-camera system to obtain 3D gait

978-1-4244-3993-5/09/$25.00 ©2009 IEEE

Authorized licensed use limited to: Univ of Calif Riverside. Downloaded on September 11, 2009 at 13:42 from IEEE Xplore. Restrictions apply.

Table 1. Summary of biometrics approaches for human recognition

Authors	Technique	Data	Biometrics
Zhao *et al.* [3]	Fitting 3D body model to video sequences captured by multi-camera to extract the lengths of key segments and the motion trajectory of lower limbs	2D	Gait
Huang *et al.* [4]	Combination of the results of different views into a common distance metric for the evaluation of similarity	2D	Gait
Seely *et al.* [5]	Classification by non-normalized average silhouettes from three orthogonal viewpoints; side-on, front-on, and top-down view	2D	Gait
Kakadiaris *et al.* [9]	Annotated face model fitting and its deformed model geometry analysis by wavelet transform	3D	Face
Malassiotis *et al.* [10]	Similarity measurement computed by finger width and curvature measurements sampled by finger length	3D	Finger
Yan and Bowyer [11]	Modified ICP-based shape matching using ear point-cloud representation to align probe surface with gallery surface	3D	Ear
Bhanu and Chen [12]	Complete human identification system consisting of 3D ear detection, 3D ear identification/verification, and performance prediction	3D	Ear
Tsalakanidou *et al.* [13]	Combined face classifier based on probabilistic matching algorithm and finger classifier based on similarity	3D	Face & Finger
Theoharis *et al.* [14]	Wavelet transform for geometries derived from annotated face model and annotated ear model	3D	Face & Ear
This paper	Score matching for dynamic and static features of gait sequence recovered by interpolation of representative poses	3D	Gait

biometrics, the low resolution and the small number of measurement points are not well suited and qualified to make precise models of human body in motion. In contrast, a projector camera system, used in this paper, captures whole human body data with 3 mm depth resolution. The number of measurement points on the entire human body is approximately one million. The data rates are $\sim$2-3 sec for the 3D data on the entire human body. To our knowledge there is no publication in the CVPR field that has been able to generate 3D data of (markerless) walking humans from a multi-camera passive stereo with such a high accuracy and data rates. Since the data is dense and it is at a high resolution, we can interpolate the data to fill-in between gait acquisitions for improved human motion analysis. Therefore, the projector-camera system used in this paper is well suited for gait recognition and it provides a firm basis for systematic evaluation of gait biometrics in 3D.

In this paper, we present a recognition method using 3D gait biometrics from a projector-camera system. 3D human body data consisting of representative poses over one gait cycle are captured. 3D human body model is fitted to the body data using a bottom-up approach. The entire gait sequence is recovered in 3D from the fitted 3D body model. Then, gait features are defined by dynamic and static features. The similarity measure based on gait features is used for recognition.

2. Related Work and Our Contributions

Biometrics systems generally use 2D images and videos, and extract features for human recognition. They are successfully gaining ground and are available for security applications. Biometrics modalities with depth information are an attractive resource. The following are representative recognition approaches for 3D biometrics.

Kakadiaris *et al.* [9] developed a fully automatic 3D face recognition system based on a composite alignment algorithm to register 3D facial scans with a 3D facial model. The geometry image and normal map image are created by the deformed facial model. They are analyzed by a wavelet transform and the coefficients are used for authentication. Malassiotis *et al.* [10] proposed an authentication system based on measurements of 3D finger geometry using a real-time and low-cost 3D sensor. The similarity between training and testing set are computed by the finger width and curvature measurements sampled by the finger length. Yan and Bowyer [11] proposed a complete system for ear biometrics including automated ear region segmentation and 3D ear shape matching for recognition. The system uses an improved interactive closest point (ICP) method, combined with point-to-point and point-to-surface approaches, to align a probe surface with a gallery surface. Multimodal 3D biometrics approaches have been developed in recent years. Tsalakanidou *et al.* [13] presented an authentication system based on the fusion of 3D face and finger biometrics. Theoharis *et al.* [14] presented a unified approach to fuse 3D facial and ear data. These methods achieve high recognition rate when compared to a single modality approach. As compared to all the work presented in Table 1, there are various approaches for 2D and 3D biometrics. While biometrics approaches using 3D face, finger, and ear data

Authorized licensed use limited to: Univ of Calif Riverside. Downloaded on September 11, 2009 at 13:42 from IEEE Xplore. Restrictions apply.

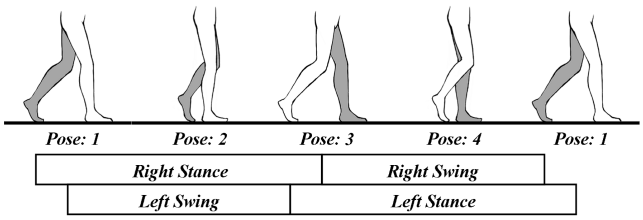


Figure 1. Gait cycle expressed by swing phase and stance phase.

have been proposed, gait recognition method still utilizes 2D data. Therefore, we attempt to tackle human recognition using 3D gait biometrics where both the modeling and the test data are obtained in 3D.

The contributions of this paper are (a) Representative poses of walking human are captured in 3D by a novel projector-camera system unlike motion capture methods with/without markers using a multi-camera system. (b) Continuous motion is recovered by interpolation of the representative poses whose joint positions are estimated by a bottom-up approach using an articulated limb model. (c) 3D gait biometrics based on interpolated motion provides highly reliable authentication.

3. Technical Approach

In our approach representative poses of walking human are captured by using a high resolution projector-camera system. Then, we fit 3D human body model to the captured 3D data in the following order: torso, limb, and head. Next, gait sequence is recovered by interpolation of the fitted body models to the real data. Then gait features as defined by dynamic and static features are obtained. Finally, feature matching is performed based on the similarity measure for individual human recognition.

3.1. 3D Human Body Data

Gait has two distinct periods: a swing phase, when the foot does not touch the ground moving the leg forward, and a stance phase, when the foot touches the ground. Murray *et al.* [15] propose that a gait cycle is the time interval between instances of initial foot-to-floor contact, called *heel strike*, for the same foot.

Fig. 1 is the gait cycle expressed by the swing phase and the stance phase. The cycle begins with *foot touch* which marks the start of the swing phase. The body weight is transferred onto the other leg and the leg swings forward to meet the ground in front of the other foot. The cycle ends with the *foot touch*. The start of stance phase is when the heel strikes the ground. The ankle flexes to bring the foot flat on the ground and the body weight transferred onto it. The end of stance phase is when the heel leaves the ground.

We measure four poses during the cycle by a projector-camera system. The projector-camera system captures high resolution and highly accurate whole human body data. It includes approximately one million 3D points, **x**, in a few (2-3) seconds. A subject has the following posture conditions:

1. Right foot touches the ground. Right leg (left hand) is in front of the torso and left leg (right hand) is at the back of the torso. The length of stride is the longest during walking.

2. Right foot touches the ground and left foot leaves the ground. Right leg is vertical to the ground and left leg is at the back of the torso. Both hands are along the sides.

3. Left foot touches the ground. Left leg (right hand) is in front of the torso and right leg (left hand) is at the back of the torso. The length of stride is the longest during walking.

4. Left foot touches the ground and right foot leaves the ground. Left leg is vertical to the ground and right leg is at the back of the torso. Both hands are along the sides.

Currently, gait databases have tens of images during a gait cycle [16, 17]. In this paper we assumed that the measured poses are four of them.

3.2. 3D Human Body Model

The model of the human body is based on a kinematic tree consisting of 12 segments, as illustrated in Fig. 2. Each body segment, i, is approximated by a 3D tapered cylinder which has one free parameter, l^i: the cylinder length. It has two degrees of freedom rotational joint, $[\theta_x^i, \theta_z^i]$, in the local coordinate system (O_i-X_i-Y_i-Z_i). Upper torso is the root segment, i.e., the parent of lower torso, right upper leg, and left upper leg. Similarly, other segments are linked to parent segments by the rotational joints.

The articulated structure of the human body has a total of 40 degrees of freedom (DOFs). The pose is described by a 6-D vector, **p**, representing global position and rotation, a 22-D vector, **q**, representing the joint angles, and a 12-D vector, **r**, representing the lengths of body part as follows.

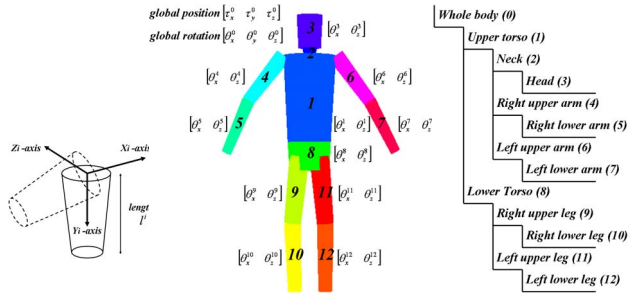
$$\mathbf{p} = [\tau_x^0, \tau_y^0, \tau_z^0, \theta_x^0, \theta_y^0, \theta_z^0] \quad (1)$$

$$\mathbf{q} = [\theta_x^1, \theta_z^1, \theta_x^3, \theta_z^3, \theta_x^4, \theta_z^4, \theta_x^5, \theta_z^5, \theta_x^6, \theta_z^6, \theta_x^7, \theta_z^7, \theta_x^8, \theta_z^8, \theta_x^9, \theta_z^9, \theta_x^{10}, \theta_z^{10}, \theta_x^{11}, \theta_z^{11}, \theta_x^{12}, \theta_z^{12}] \quad (2)$$

$$\mathbf{r} = [l^1, l^2, l^3, l^4, l^5, l^6, l^7, l^8, l^9, l^{10}, l^{11}, l^{12}] \quad (3)$$

Here, neck is the fixed segment between head and upper torso, so that we do not consider neck angles. The combination of the representative four poses is denoted by u. Joint

Authorized licensed use limited to: Univ of Calif Riverside. Downloaded on September 11, 2009 at 13:42 from IEEE Xplore. Restrictions apply.



(a) (b) (c)

Figure 2. 3D human body model. (a) Tapered cylinder with two angular DOF. (b) Body model approximated by 12 segments. (c) Hierarchical structure.

DOF values concatenated along the kinematic tree define the kinematic pose, $\mathbf{k}$, as a tuple, $[\mathbf{p}, \mathbf{q}, \mathbf{r}, u]$, where $\mathbf{p} \in \mathbb{R}^6$, $\mathbf{q} \in \mathbb{R}^{22}$, $\mathbf{r} \in \mathbb{R}^{12}$, $u = \{u_1, u_2, u_3, u_4\}$.

In the previous works, segments are linked to parent segments by either 1-DOF (hinge), 2-DOF (saddle) or 3-DOF (ball and socket) rotational joints [18]. We use only 2-DOF rotational joints, because the 3D tapered cylinder has rotational symmetry along the direction orthogonal to the radial direction. Therefore, we eliminate unnecessary variables.

3.3. Model Fitting

Let us consider human body modeling and its problems. Modeling methods, which use *ideal data*, sometimes fail when applied to real data [19]. The real data captured by projector-camera systems have some problems. For example, the projector-camera system cannot cover well particular body parts, such as the groin region, axillary region, and side of a human body, so that 3D points of the real data are not independently and identically distributed [20]. In addition, the body sways and deep color clothes also have detrimental effects such as holes and gaps.

In this section, a modeling method for dealing with the problems occurring in real data is proposed. Our approach to modeling a walking human incorporates four separate steps: body axes estimation, torso detection, arms/legs detection, and head/neck detection.

3.3.1 Body Axes

The intuition behind the principal component analysis (PCA) is to find a set of base vectors, so that they explain the maximum amount of variance of the data [21]. PCA is applied to determine coronal axis (X-axis), vertical axis (Y-axis), and sagittal axis (Z-axis). The centroid of a human body and the three axes are shown in Fig. 3.

Our approach to determining the three axes and the centroid incorporates two separate steps. First, we compute the eigenvectors and the mean vector using the data of the whole human body. The first eigenvector, $\mathbf{e}_1$, and the mean vector, $\mathbf{m}$, define the vertical axis and the centroid. The data of arms and legs do not affect the estimation of the vertical axis and the centroid adversely because they are at symmetric positions in a horizontal direction. Second, we compute the eigenvectors using the extracted torso data (after torso detection, see next subsection **3.3.2**). The second eigenvector, $\mathbf{e'}_2$, and the third eigenvector, $\mathbf{e'}_3$, define the coronal axis and the sagittal axis, respectively. The torso data is convex and has symmetrical shape even if a subject is walking, so that the two axes can be estimated robustly. Finally, the world coordinate system (O-X-Y-Z) is defined by associating the coronal axis, the vertical axis, the sagittal axis, and the centroid with X-axis, Y-axis, Z-axis, and the origin O.

3.3.2 Torso

We use the cross-section of a human body to detect upper torso and lower torso. Body data have some holes and gaps, so that the sampling and interpolation process for the extraction of cross-section are required.

First, the cross-section is divided into R regions radiating from a 3D point, $\mathbf{c}_h$, which is the intersection of the Y-axis, with the X-Z plane. Here, the index h corresponds to the body height. For each region, the closest point from the intersection point is left and the others are removed. If there are no points, one linear interpolated point is calculated using the neighbors. Next, the sample points are connected by the line segments. The area inside the line segments defined by s_i are used for the detection of upper torso and lower torso.

The height of the centroid is denoted by g. We assume the cross-sectional area, s_g, is the boundary between the upper torso and lower torso. The cross-sectional area of the top of upper torso, s_m, and the cross-sectional area of the base of lower torso, s_n, are given by

$$s_m = \delta_{ut} s_g \quad (4)$$

$$s_n = \delta_{lt} s_g \quad (5)$$

where δ_{ut} and δ_{lt} are weight parameters. We compute all the cross-sectional areas, and then search for the similar values in a vertical direction. When s_i is smaller than s_m (or s_n), i is the height of the top of the upper torso (or the base of the lower torso). Then, directional vectors, $\mathbf{n}_{ut}$ and $\mathbf{n}_{lt}$, of upper torso and lower torso are estimated as

$$\mathbf{n}_{ut} = \text{median}\left(\frac{\mathbf{c}_h - \mathbf{c}_g}{\|\mathbf{c}_h - \mathbf{c}_g\|}\right) \qquad g < h \leq m \quad (6)$$

$$\mathbf{n}_{lt} = \text{median}\left(\frac{\mathbf{c}_h - \mathbf{c}_g}{\|\mathbf{c}_h - \mathbf{c}_g\|}\right) \qquad n \leq h < g \quad (7)$$

Authorized licensed use limited to: Univ of Calif Riverside. Downloaded on September 11, 2009 at 13:42 from IEEE Xplore. Restrictions apply.

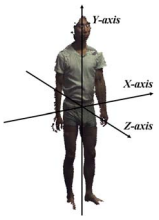


Figure 3. Body axes estimation by PCA.

The joint angles of upper torso and lower torso are obtained by the directional vectors. One is the angle between $\mathbf{n}_{ut}$ (or $\mathbf{n}_{lt}$) and the X-axis, and the other is the angle between $\mathbf{n}_{ut}$ (or $\mathbf{n}_{lt}$) and the Z-axis. Therefore, tapered cylinders can be fitted along the directional vectors.

3.3.3 Arms and Legs

We use the fitting of tapered cylinders to detect arms and legs. In this subsection, right/left-upper/lower-arm and right/left-upper/lower-leg are detected by using the same method. Thus, the detection of two of them, which are called the *upper part* and the *lower part* (e.g. right upper arm and right lower arm), takes place.

The line segments between the top and base of the cylinders of the upper part and lower part are defined as $L_{up} = \{\mathbf{o}_{up} + \lambda \mathbf{n}_{up} \mid \lambda \in \mathbb{R}^3\}$ and $L_{lp} = \{\mathbf{o}_{lp} + \lambda \mathbf{n}_{lp} \mid \lambda \in \mathbb{R}^3\}$. Here, $\mathbf{o}_{up}$ and $\mathbf{o}_{lp}$ are joint points (shoulder and elbow, or hip and knee). $\mathbf{n}_{up}$ and $\mathbf{n}_{lp}$ are the corresponding directional vectors, which are used as $\mathbf{n}_{rua}$ and $\mathbf{n}_{rla}$ for the right arm, and $\mathbf{n}_{rul}$ and $\mathbf{n}_{rll}$ for the right leg. Therefore, the distance between the line and 3D points can be written as

$$d_{up} = \sum_{\mathbf{x}_{up} \in upper\ part(\mathbf{x})} \|\mathbf{x}_{up} - \mathbf{o}_{up} + \frac{\mathbf{x}_{up}\mathbf{n}_{up} - \mathbf{o}_{up}\mathbf{n}_{up}}{\mathbf{n}_{up}\mathbf{n}_{up}}\mathbf{n}_{up}\|^2 \quad (8)$$

$$d_{lp} = \sum_{\mathbf{x}_{lp} \in lower\ part(\mathbf{x})} \|\mathbf{x}_{lp} - \mathbf{o}_{lp} + \frac{\mathbf{x}_{lp}\mathbf{n}_{lp} - \mathbf{o}_{lp}\mathbf{n}_{lp}}{\mathbf{n}_{lp}\mathbf{n}_{lp}}\mathbf{n}_{lp}\|^2 \quad (9)$$

where $\mathbf{x}_{up}$ and $\mathbf{x}_{lp}$ are 3D points within each of the parts. We first seek the direction vectors to minimize each of the functions, and then two sets of two joint angles are estimated from the directional vectors. Accordingly, tapered cylinders can be fitted to the arms and legs along the directional vectors.

3.3.4 Head and Neck

Let us consider the difference between head and the other body parts. In the measurement of the head, it is sometimes difficult to capture the shape of hair on the head, because of the low sensitivities to deep color. Therefore, face shape and neck shape are used for the detection.

First, the sizes of head and neck are estimated from the distribution of 3D points in the X-Y plane and Y-Z plane. Next, the directional vector of head, denoted by $\mathbf{n}_h$, is determined, and then the tapered cylinder is fitted to the head. As stated above, neck does not rotate in our model independently, so that the neck's tapered cylinder is placed on upper torso.

3.4. Gait Reconstruction

Gait sequence composed of tens or hundreds of poses is required to analyze and recognize. The representative four poses obtained by fitting body models to body data are used to recover the other poses.

Assuming that the motion between pose α and pose β varies linearly, kinematic pose, $\mathbf{k}_f = [\mathbf{p}_f, \mathbf{q}_f, \mathbf{r}_f, u]$, at frame f ($\alpha < f < \beta$) can be written as

$$\mathbf{p}_f = \mathbf{p}_\alpha + (f - \alpha)\mathbf{v} \quad (10)$$

$$\mathbf{q}_f = \mathbf{q}_\alpha + \frac{f - \alpha}{\beta - \alpha}(\mathbf{q}_\beta - \mathbf{q}_\alpha) \quad (11)$$

$$\mathbf{r}_f = (\mathbf{r}_{u_1} + \mathbf{r}_{u_2} + \mathbf{r}_{u_3} + \mathbf{r}_{u_4})/4 \quad (12)$$

where $\mathbf{v}$ is velocity vector which includes speed and direction, and combination of α and β is expressed by $\{\alpha, \beta\} \in \{\{u_1, u_2\}, \{u_2, u_3\}, \{u_3, u_4\}, \{u_4, u_1\}\}$. The equations allow interpolation of joint angles and lengths of body part. Therefore, arbitrary poses between representative poses can be recovered.

3.5. Feature Matching

Gait features are divided into two types: (a) dynamic features and (b) static features. For example, the length of stride is one of significant features of human gait. It can be computed by the leg length and its varying angles between poses. In addition, all of joint positions can be computed by using the same method. Therefore, both dynamic feature and static feature are used for recognition.

We define the dynamic feature as joint angles, $\mathbf{q}_{i,j}$, and static feature as lengths of the body parts, $\mathbf{r}_{i,j}$. Here, i is personal identification number, and j is a pose index. To transform these values into a common domain, the normalization is given by

$$\mathbf{q}'_{i,j} = \frac{\mathbf{q}_{i,j} - \boldsymbol{\mu}_q}{\boldsymbol{\sigma}_q} \quad (13)$$

$$\mathbf{r}'_{i,j} = \frac{\mathbf{r}_{i,j} - \boldsymbol{\mu}_r}{\boldsymbol{\sigma}_r} \quad (14)$$

where $\boldsymbol{\mu}_q = \frac{1}{M}\frac{1}{N}\sum_i^M \sum_j^N \mathbf{q}_{i,j}$ $\boldsymbol{\mu}_r = \frac{1}{M}\frac{1}{N}\sum_i^M \sum_j^N \mathbf{r}_{i,j}$

Authorized licensed use limited to: Univ of Calif Riverside. Downloaded on September 11, 2009 at 13:42 from IEEE Xplore. Restrictions apply.

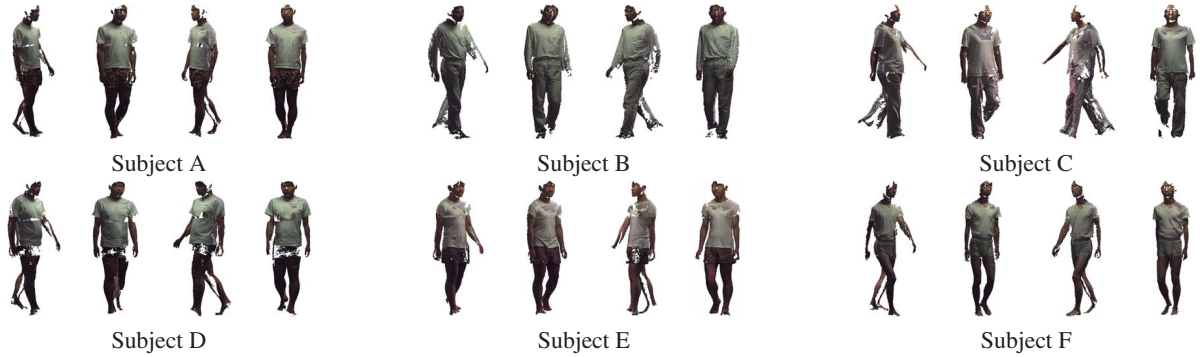


Figure 4. 3D human body data. Representative poses of walking humans.

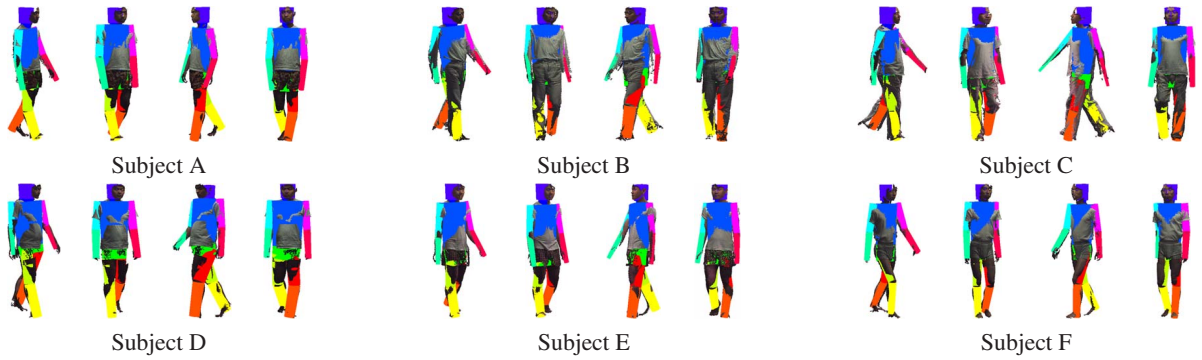


Figure 5. Fitted 3D models to four poses of gait cycle.

$$\boldsymbol{\sigma}_q = (\frac{1}{M}\frac{1}{N}\sum_i^M\sum_j^N(\mathbf{q}_{i,j}-\boldsymbol{\mu}_q)(\mathbf{q}_{i,j}-\boldsymbol{\mu}_q)^T\cdot\mathbf{I})^{1/2}\cdot\mathbf{k}$$

$$\boldsymbol{\sigma}_r = (\frac{1}{M}\frac{1}{N}\sum_i^M\sum_j^N(\mathbf{r}_{i,j}-\boldsymbol{\mu}_r)(\mathbf{r}_{i,j}-\boldsymbol{\mu}_r)^T\cdot\mathbf{I})^{1/2}\cdot\mathbf{k}$$

In the formulations, $\boldsymbol{\mu}_q$, $\boldsymbol{\mu}_r$ are arithmetic means of dynamic and static features, and $\boldsymbol{\sigma}_q$, $\boldsymbol{\sigma}_r$ are standard deviations of dynamic and static features, M, N are the numbers of people and poses, with the matrix $\mathbf{I} = diag(1,1,1,\ldots,1)$ and the vector $\mathbf{k} = [1,1,1,\ldots,1]$. Both features are concatenated on a feature vector $\boldsymbol{\phi}_{i,j} = [\mathbf{q}'_{i,j}, \mathbf{r}'_{i,j}]$. If dynamic feature is only used, a feature vector is defined as $\boldsymbol{\phi}_{i,j} = [\mathbf{q}'_{i,j}]$.

Suppose that unknown feature vector, $\boldsymbol{\phi}_U$, is one of $M \times N$ feature vectors, $\boldsymbol{\phi}_{i,j}$. The minimum value of matching scores can be written as

$$s = \min_{i,j} \|\boldsymbol{\phi}_U - \boldsymbol{\phi}_{i,j}\| \tag{15}$$

The matching score is computed as L_2 distance. For unknown data, the personal identification number and pose index are recognized.

4. Experimental Results

The experiments were performed on the 3D human body data set collected by a projector-camera system. It contains twenty-four body data from the representative four poses of six subjects $X \in \{A, B, C, D, E, F\}$.

4.1. Sensing and Modeling

The body data of representative poses are captured by the projector-camera system [22]. The system consisted of nine laser rangefinders, which acquires nine range images in ∼two-three seconds with 640×480 pixels, 3 mm depth resolution, and measurement accuracy within 2 mm. Data from nine different views are integrated. Fig. 4 is the measurement results of walking humans [1]. The number of measured 3D points is about 1/2 to one million depending on the subject and the pose. Fig. 5 is the results of human body modeling. The body model is fitted to the captured body data, so that their joint angles and lengths of body parts are obtained.

4.2. Gait Reconstruction

Fig. 6 is the results of gait reconstruction. We define that the one gait cycle is composed of twenty frames

[1]For the modeling in this paper we used $R = 36$, $\delta_{ut} = 0.25$, and $\delta_{lt} = 0.5$.

Authorized licensed use limited to: Univ of Calif Riverside. Downloaded on September 11, 2009 at 13:42 from IEEE Xplore. Restrictions apply.

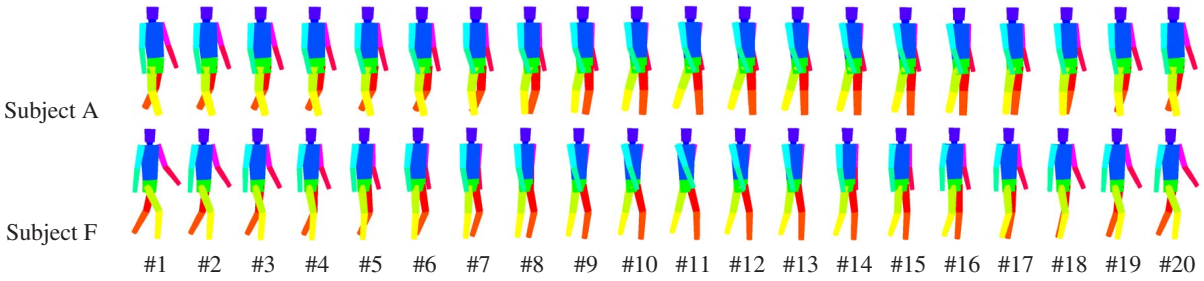


Figure 6. Gait sequence composed of twenty frames recovered by four poses.

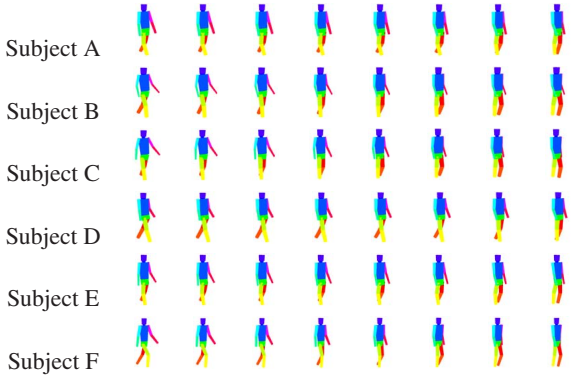


Figure 7. Examples of training data for six subjects.

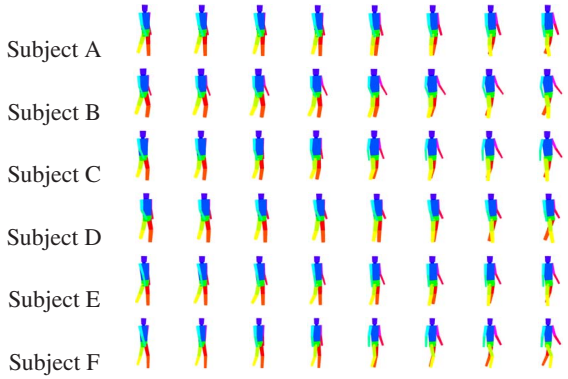


Figure 8. Examples of testing data for six subjects.

$Y \in \{1, 2, \ldots, 20\}$. The speed is given by dividing the stride length by the number of poses and the direction is given manually. Four of them are representative poses, indicated by the frame index 1, 6, 11, and 16, and the others are interpolated poses, indicated by the frame index 2-5, 7-10, 12-15, and 17-20.

4.3. Recognition

4.3.1 Training and Testing Data

The representative poses $u = \{u_1, u_2, u_3, u_4\}$ and their symmetric poses $\bar{u} = \{\bar{u}1, \bar{u}2, \bar{u}3, \bar{u}4\}$ are used for the experiment. The symmetric poses $\bar{u}1$, $\bar{u}2$, $\bar{u}3$, $\bar{u}4$ are symmetric to $u3$, $u4$, $u1$, $u2$, respectively. They are synthesized by allocating right (or left) side parameters of representative poses to left (or right) side parameters of symmetrical poses.

For the training data, two gait sequences are recovered by using two combinations of representative poses and symmetrical poses. Fig. 7 shows the training data of six subjects. One gait sequence is recovered by four poses $c1 = \{u1, \bar{u}2, u3, \bar{u}4\}$, and the other one is recovered by four poses $c2 = \{\bar{u}1, u2, \bar{u}3, u4\}$. Each subject has 40 poses, so that training data contains a total of 240 kinematic poses.

For the testing data, one gait sequence is recovered by representative poses $c3 = \{u_1, u_2, u_3, u_4\}$. Fig. 8 shows the testing data of six subjects. This sequence includes the representative four poses and sixteen interpolated poses. The sixteen interpolated poses are unique and also they are not included in the training data. Therefore, we utilize 96 kinematic poses of six subjects for testing. There is absolutely no overlap between the training and testing data.

4.3.2 Recognition

In order to evaluate the proposed method, identification rate and average pose error are obtained. The identification rate is obtained by dividing the number of recognized subject by the number of testing data. The pose error is the frame difference between the estimated pose and the ideal pose.

Table 2 shows that we achieve 98.96 percent using dynamic feature and 100.0 percent using both dynamic and static features for the identification rate. When only dynamic feature is used, the method fails to recognize testing data Subject D with pose 14 who should not be recognized as the training data for Subject B with pose 13. Although body types between two subjects are different, their joint angles, i.e. leg and arm swings, are quite similar. In contrast, we achieve 0.41 using dynamic feature and 1.31 using both features for the average pose error. The experiment using dynamic feature has acceptable results, because it focuses on estimating poses, i.e. dynamic feature cannot consider lengths of body part. Therefore, both dynamic and static features are useful for gait recognition.

Authorized licensed use limited to: Univ of Calif Riverside. Downloaded on September 11, 2009 at 13:42 from IEEE Xplore. Restrictions apply.

Table 2. Identification rate and average pose error.

Features	Identification rate (%)	Average pose error (frame)
Dynamic	98.96	0.41
Dynamic and Static	100.0	1.31

5. Conclusions

One of the significant weaknesses of current gait recognition methods using a 3D kinematic model has been lack of enough pixels on the human body to fit the 3D models accurately in the 2D video or the 3D data obtained with a multi-camera system. We proposed a new approach for biometric authentication based on 3D human gait. In our approach the body data are captured by a high performance projector-camera system. Unlike the multi-camera passive stereo systems used to date this 3D data is of high resolution and has high accuracy. As a result, the fitted 3D body models and the reconstructed synthetic poses in a gait cycle are quite accurate, including the interpolation of the joint angles and lengths of the body parts. Using training and testing experiments we verified that 3D gait biometrics provide high identification rate to recognize a human subject and his/her pose. In the future we intend to expand our database and make it publicly available.

Acknowledgements

This work was partially supported by ONR grant N00014-07-1-0931.

References

[1] Z. Liu and S. Sarkar. Improved gait recognition by gait dynamics normalization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(6):863–876, 2006.

[2] M. Goffredo, R. D. Seely, J. N. Carter, and M. S. Nixon. Markerless view independent gait analysis with self-camera calibration. In *Proceedings of International Conference on Automatic Face and Gesture Recognition*, 2008.

[3] G. Zhao, G. Liu, H. Li, and M. Pietikäinen. 3D gait recognition using multiple cameras. In *Proceedings of International Conference on Automatic Face and Gesture Recognition*, 529–534, 2006.

[4] X. Huang and N. V. Boulgouris. Human gait recognition based on multiview gait sequences. *EURASIP Journal on Advances in Signal Processing*, 2008:1–8, 2008.

[5] R. D. Seely, S. Samangooei, L. Middleton, J. N. Carter, and M. S. Nixon. The university of southampton multi-biometric tunnel and introducing a novel 3D gait dataset. In *Proceedings of International Conference on Biometrics: Theory, Applications and Systems*, 1–6, 2008.

[6] K. M. Cheung, S. Baker, and T. Kanade. Shape-from-silhouette across time part Ⅱ: applications to human modeling and markerless motion tracking. *International Journal of Computer Vision*, 63(3):225–245, 2005.

[7] F. Caillette, A. Galata, and T. Howard. Real-time 3-D human body tracking using learnt models of behavior. *Computer Vision and Image Understanding*, 109(2):112–125, 2008.

[8] K. M. Cheung, S. Baker, and T. Kanade. Shape-from-silhouette across time part I: theory and algorithms. *International Journal of Computer Vision*, 63(3):221–247, 2005.

[9] I. A. Kakadiaris, G. Passalis, G. Toderici, M. N. Murtuza, Y. Lu, N. Karampatziakis, and T. Theoharis. Three-dimensional face recognition in the presence of facial expressions: an annotated deformable model approach. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 29(4):640–649, 2007.

[10] S. Malassiotis, N. Aifanti, and M. G. Strintzis. Personal authentication using 3-D finger geometry. *IEEE Transactions on Information Forensics and Security*, 1(1):12–21, 2006.

[11] P. Yan and K. W. Bowyer. Biometric recognition using 3D ear shape. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(8):1297–1308, 2007.

[12] B. Bhanu and H. Chen. *Human ear recognition by computer*. Springer-Verlag, New York, 2008.

[13] F. Tsalakanidou, S. Malassiotis, and M. G. Strintzis. A 3D face and hand biometric system for robust user-friendly authentication. *Pattern Recognition Letters*, 28(16):2238–2249, 2007.

[14] T. Theoharis, G. Passalis, G. Toderici, and I. A. Kakadiaris. Unified 3D face and ear recognition using wavelets on geometry images. *Pattern Recognition*, 41(3):796–804, 2008.

[15] M. P. Murray, A. B. Drought, and R. C. Kory. Walking patterns of normal men. *Journal of Bone and Joint Surgery*, 46A(2):335–360, 1964.

[16] R. Gross and J. Shi. The CMU motion of body (MoBo) database. Technical Report CMU-RI-TR-01-18, Robotics Institute, Carnegie Mellon University, 2001.

[17] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, and K. W. Bowyer. The humanID gait challenge problem: data sets, performance, and analysis. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 27(2):162–177, 2005.

[18] M. Vondrak, L. Signal, and O. C. Jenkins. Physical simulation for probabilistic motion tracking. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1–8, 2008.

[19] H. Yu, S. Qin, D. K. Wight, and J. Kang. Generation of 3D human models with different levels of detail through point-based simplification. In *Proceedings of International Conference on "Computer as a Tool"*, 1982–1986, 2007.

[20] N. Werghi. Segmentation and modeling of full human body shape from 3-D scan data: a survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, 37(6):1122–1136, 2007.

[21] I. T. Jolliffe. *Principal Component Analysis – 2nd Edition*. Springer-Verlag, New York, 2002.

[22] K. Yamauchi and Y. Sato. 3D human body measurement by multiple range images. In *Proceedings of International Conference on Pattern Recognition*, 4:833–836, 2006.

Authorized licensed use limited to: Univ of Calif Riverside. Downloaded on September 11, 2009 at 13:42 from IEEE Xplore. Restrictions apply.