LETTER Communicated by Yaser Yacoob

3D Periodic Human Motion Reconstruction from 2D Motion Sequences

Zonghua Zhang
z.zhang@hw.ac.uk
Nikolaus F. Troje
troje@post.queensu.ca
BioMotionLab, Department of Psychology, Queen's University, Kingston, Ontario K7M 3N6 Canada

We present and evaluate a method of reconstructing three-dimensional (3D) periodic human motion from two-dimensional (2D) motion sequences. Using Fourier decomposition, we construct a compact representation for periodic human motion. A low-dimensional linear motion model is learned from a training set of 3D Fourier representations by means of principal components analysis. Two-dimensional test data are projected onto this model with two approaches: least-square minimization and calculation of a maximum a posteriori probability using the Bayes′ rule. We present two different experiments in which both approaches are applied to 2D data obtained from 3D walking sequences projected onto a plane. In the first experiment, we assume the viewpoint is known. In the second experiment, the horizontal viewpoint is unknown and is recovered from the 2D motion data. The results demonstrate that by using the linear model, not only can missing motion data be reconstructed, but unknown view angles for 2D test data can also be retrieved.

1 Introduction

Human motion contains a wealth of information about the actions, intentions, emotions, and personality traits of a person, and human motion analysis has widespread applications—in surveillance, computer games, sports, rehabilitation, and biomechanics. Since the human body is a complex articulated geometry overlaid with deformable tissues and skin, motion analysis is a challenging problem for artificial vision systems. General surveys of the many studies on human motion analysis can be found in recent review articles (Gavrila, 1999; Aggarwal & Cai, 1999; Buxton, 2003; Wang, Hu, & Tan, 2003; Moeslund & Granum, 2001; Aggarwal, 2003; Dariush, 2003). The existing approaches to human motion analysis can be roughly divided into two categories: model-based methods and model-free methods. In the model-based methods, an a priori human model is used to represent the observed subjects; in model-free methods, the motion information is derived directly from a sequence of images. The main drawback of model-free methods is that they are usually designed to work with images taken from a known viewpoint. Model-based approaches support viewpoint-independent processing and have the potential to generalize across multiple viewpoints (Moeslund & Granum, 2001; Cunado, Nixon, & Carter, 2003; Wang, Tan, Ning, & Hu, 2003; Jepson, Fleet, & El-Maraghi, 2003; Ning, Tan, Wang, & Hu, 2004).

Most of the existing research (Gavrila, 1999; Aggarwal & Cai, 1999; Buxton, 2003; Wang, Hu, & Tan, 2003; Moeslund & Granum, 2001; Aggarwal, 2003; Dariush, 2003; Cunado et al., 2003; Wang, Tan et al., 2003; Jepson et al., 2003; Ning et al., 2004) has been focused on the problem of tracking and recognizing human activities through motion sequences. In this context, the problem of reconstructing 3D human motion from 2D motion sequences has received increasing attention (Rosales, Siddiqui, Alon, & Sclaroff, 2001; Sminchisescu & Triggs, 2003; Urtasun & Fua, 2004a, 2004b; Bowden, 2000; Bowden, Mitchell, & Sarhadi, 2000; Ong & Gong, 2002; Yacoob & Black, 1999).

The importance of 3D motion reconstruction stems from applications such as surveillance and monitoring, human body animation, and 3D human-computer interaction. Unlike 2D motion, which is highly view dependent, 3D human motion can provide robust recognition and identification. However, existing systems need multiple cameras to get 3D motion information. If 3D motion can be reconstructed from a single camera viewpoint, there are many potential applications. For instance, using 3D motion reconstruction, one could create a virtual actor from archival footage of a movie star, a difficult task for even the most skilled modelers and animators. 3D motion reconstruction could be used to track human body activities in real time (Arikan & Forsyth, 2002; Grochow, Martin, Hertzmann, & Popović, 2004; Chai & Hodgins, 2005; Aggarwal & Triggs, 2004; Yacoob & Black, 1999). Such a system may be used as a new and effective human-computer interface for virtual reality applications.

In the model of Kakadiaris and Metaxas (2000), motion estimation of human movement was obtained from multiple cameras. Bowden and colleagues (Bowden, 2000; Bowden et al., 2000) used a statistical model to reconstruct 3D postures from monocular image sequences. Just as a 3D face can be reconstructed from a single image using a morphable model (Blanz & Vetter, 2003), they reconstructed the 3D structure of a subject from a single view of its outline. Ong and Gong (2002) discussed three main issues in the linear combination method: choosing the examples to build a model, learning the spatiotemporal constraints on the coefficients, and estimating the coefficients. They applied their method to track moving 3D skeletons of humans. These models (Bowden, 2000; Bowden et al., 2000; Ong & Gong, 2002) are based on separate poses and do not use the temporal information that connects them. The result is a series of reconstructed 3D human postures.

Using principal components analysis (PCA), Yacoob and Black (1999) built a parameterized model for image sequences to model and recognize activity. Urtasun and Fua (2004a) presented a motion model to track the human body and then (Urtasun & Fua, 2004b) extended it to characterize and recognize people by their activities. Leventon and Freeman (1998) and Howe, Leventon, and Freeman (1999) studied reconstruction of human motion from image sequences using Bayes' rule, which is in many respects similar to our approach. However, they did not present a quantitative evaluation of the 3D motion reconstructions corresponding to the missing dimension. For viewpoint reconstruction from motion data, some researchers (Giese & Poggio, 2000; Agarwal & Triggs, 2004; Ren, Shakhnarovich, Hodgins, Pfister, & Viola, 2005) quantitatively evaluated the performance for their motion model.

Incorporating temporal data into model-based methods requires correspondence-based representations, which separate the overall information into range-specific information and domain-specific information (Ramsay & Silverman, 1997; Troje, 2002a). In the case of biological motion data, range-specific information refers to the state of the actor at a given time in terms of the location of a number of feature points. Domain-specific information refers to when a given position occurs. Yacoob and Black (1999) addressed temporal correspondence in their walking data by assuming that all examples had been temporally aligned. Urtasun and Fua (2004a, 2004b) chose one walking cycle with the same number of samples. Giese and Poggio (2000) presented a learning-based approach for the representation of complex motion patterns based on linear combinations of prototypical motion sequences. Troje (2002a) developed a framework that transformed biological motion data into a linear representation using PCA. This representation was used to construct a sex classifier with a reasonable classification performance. These authors (Troje, 2002a; Giese & Poggio, 2000) pointed out that finding spatiotemporal correspondences between motion sequences is the key issue for the development of efficient models that perform well on reconstruction, recognition, and classification tasks.

Establishing spatiotemporal correspondences and using them to register the data to a common prototype is a prerequisite for designing a generative linear motion model. Our input data consist of the motion trajectories of discrete marker points, that is, spatial correspondence is basically solved. Since we are working with periodic movements—normal walking—temporal correspondence can be established by a simple linear time warp, which is defined in terms of the frequency and the phase of the walking data. This simple linear warping function can get much more complex when dealing with nonperiodic movement, and suggestions on how to expand our approach to other movements are discussed below.

In this letter, human walking is chosen as an example to study 3D periodic motion reconstruction from 2D motion sequences. On the one hand, locomotion patterns such as walking and running are periodic and highly stereotyped. On the other hand, walking contains information about the individual actor. Keeping the body upright and balanced during locomotion takes a high level of interaction of the central nervous system, sensory system (including proprioceptive, vestibular, and visual systems), and motor control systems. The solutions to the problem of generating a stable gait depend on the masses and dimensions of the particular body and its parts and are therefore highly individualized. Therefore, human walking is characterized not only by abundant similarities but also by stylistic variations. Given a set of walking data represented in terms of a motion model, the similarities are represented by the average motion pattern, while the variations are expressed in terms of the covariance matrix. Principal component analysis can be used to find a low-dimensional, orthonormal basis system that would efficiently span a motion space. Individual walking patterns are approximated in terms of this orthonormal basis system. Here, we use PCA to create a representation that is able to capture the redundancy in gait patterns in an efficient and compact way, and we test the resulting model's ability to reconstruct missing parts of the full representation.

The vision problem is not the primary purpose of this letter, and we make two assumptions to stay focused on the problem of 3D reconstruction. First, we assume that the problem of tracking feature points in the 2D image plane is solved. We represent human motion in terms of trajectories of a series of discrete markers located at the main joints of the body. Figure 1 illustrates the locations of these markers. Second, 2D motion sequences are assumed to be orthographic projections of 3D motion. Both assumptions are not critical to the general idea outlined here. In particular, the tracking problem can be treated completely independently, and there are many researchers working on this problem (Jepson et al., 2003; Ning et al., 2004; Urtasun & Fua, 2004a, 2004b; Karaulova, Hall, & Marshall, 2002).

We systematically develop a theory for recovering 3D walking data from 2D motion sequences and evaluate it using a cross-validation procedure. Single 2D test samples are generated by orthographically projecting a walker from our 3D database. We then construct a model based on the remaining walkers and project the test set onto the model. The resulting 3D reconstruction is then evaluated by comparing it to the original 3D data of this walker. This procedure is iterated through the whole data set.

Section 2 briefly describes data acquisition with a marker-based motion capture system. The details of the linear motion model and 3D motion reconstruction are given in section 3. We run our algorithm and evaluate the reconstructions in section 4. Finally, conclusions and possible future extensions are discussed in section 5.

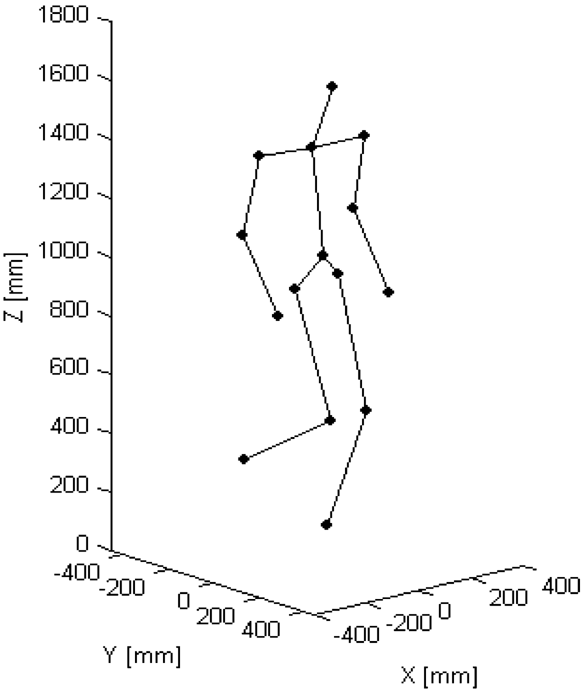

Figure 1: Human motion representation by joints. The 15 virtual markers are located at the major joints of the body (shoulders, elbows, wrists, hips, knees, ankles), the sternum, the center of the pelvis, and the center of head.

2 Walking Data Acquisition

Eighty participants—32 males and 48 females—served as subjects to acquire walking data. Their ages ranged from 13 to 59 years, and the average age was 26 years. Participants wore swimming suits, and a set of 41 retroreflective markers was attached to their bodies. Participants were requested to walk on a treadmill, and they adjusted the speed of the belt to the rate that felt most comfortable. The 3D trajectories of the markers were recorded using an optical motion capture system (Vicon; Oxford Metrics, Oxford, UK) equipped with nine CCD high-speed cameras. The system tracked the markers with a submillimeter spatial resolution and a sampling rate of 120 Hz. From the trajectories of these markers, we computed the location of 15 virtual markers according to a standard biomechanical model (BodyBuilder, Oxford Metrics), as illustrated in Figure 1. The virtual markers were located at the major joints of the body (shoulders, elbows, wrists, hips, knees, ankles), the sternum, the center of the pelvis, and the center of head (for more details, see Troje, 2002a).

3 Motion Reconstruction

In our 3D motion reconstruction method, first a linear motion model is constructed from Fourier representations of human examples by PCA, and then the missing motion data are reconstructed from the linear motion model based on two approaches: least-square minimization by pseudoinverse and calculation of a maximum a posteriori probability (MAP) by using Bayes' rule.

3.1 Linear Motion Model. The collected walking data can be regarded as a set of time series of postures $p_r(t) : t = 1, 2, \ldots, T_r, r = 1, 2, \ldots, R$, represented by the major joints, where R is the number of walkers and T_r is the number of sampled postures for walker r. Because each joint has three coordinates, the representation of a posture $p_r(t)$ consisting of 15 joints is a 45-dimensional vector.

Human walking can be efficiently described in terms of low-order Fourier series (Unuma, Anjyo, & Takeuchi, 1995; Troje, 2002b). A compact representation for a particular walker consists of the average posture (p_0), the characteristic postures of the fundamental frequency (p_1, q_1) and the second harmonic (p_2, q_2) of a discrete Fourier expansion, and the fundamental frequency (ω) to characterize this walker:

$$p(t) = p_0 + p_1 \sin(\omega t) + q_1 \cos(\omega t) + p_2 \sin(2\omega t) + q_2 \cos(2\omega t) + err. \tag{3.1}$$

The power carried by the residual term *err* is less than 3% of the overall power of the input data, and we discard it from further computations. Since the average posture and each of the characteristic postures are 45-dimensional vectors, the dimensionality of a 3D Fourier representation at this stage is $45 * 5 = 225$.

For each specific walking pattern $p_r(t)$, we can get a 3D Fourier representation w_r:

$$w_r = (p_{0,r}, p_{1,r}, q_{1,r}, p_{2,r}, q_{2,r}). \tag{3.2}$$

The advantage of Fourier representation is that it allows us to apply linear operations and easily determine temporal correspondence. Temporal information is included in the frequency and phase. After computing the frequency and phase of a walking series by Fourier analysis, we can represent the walking series with zero phase and frequency-independent characteristic postures, as shown in equation 3.2.

Every 3D Fourier representation of a walker can be treated as a point in a 225-dimensional linear space. PCA is applied to all the Fourier representations in order to learn the principal motion variations and reduce dimensionality further. To do so, all of the Fourier representations are concatenated into a matrix W, with each column containing one walker w_r, comprising the parameters $p_{0,r}$, $p_{1,r}$, $q_{1,r}$, $p_{2,r}$, and $q_{2,r}$ stacked vertically. Computing PCA on the matrix W results in a decomposition of each parameter set w of a walker into an average walker $\overline{w}$ and a series of orthogonal eigenwalkers $e_1, \ldots, e_N$:

$$w = \overline{w} + \sum_{n=1}^{N} k_n e_n, \tag{3.3}$$

$\overline{w} = (1/R) \sum w_r$ denotes the average value of all the R columns, and k_n is the projection onto eigenwalker e_n. A linear motion model is spanned by the first N eigenwalkers $e_1, \ldots, e_N$, which represent the principal variations. The exact dimensionality N of the model depends on the required accuracy of the approximation but will generally be much smaller than R. The Fourier representation of a walker can now be represented as a linear combination of eigenwalkers by using the obtained coefficients $k_1, \ldots, k_N$.

3.2 Reconstruction. We denote a 2D motion sequence as $\hat{p}(t) : t = 1, 2, \ldots, T$, which is represented in terms of its discrete Fourier components:

$$\hat{w} = (\hat{p}_0, \hat{p}_1, \hat{q}_1, \hat{p}_2, \hat{q}_2). \tag{3.4}$$

The average posture $\hat{p}_0$ and the characteristic postures $\hat{p}_1$, $\hat{q}_1$, $\hat{p}_2$, and $\hat{q}_2$ contain only 2D joint position information. These postures are concatenated into a column vector $\hat{w}$ with $30 * 5 = 150$ entries. We call this a 2D Fourier representation.

Supposing a projection matrix C relates the 2D Fourier representation and its 3D Fourier representation, reconstructing the full 3D motion means finding the right solution w to the equation

$$\hat{w} = Cw, \tag{3.5}$$

where $C : \Re^{225} \mapsto \Re^{150}$ is the projection matrix. At this point we assume C is known. Later we consider C a function of an unknown horizontal viewpoint. Equation 3.5 is an underdetermined equation system and can be solved only if additional constraints can be formulated. We constrain the possible solution to be a linear combination of eigenwalkers in the motion model outlined in the previous section, so the problem is how to calculate a set of coefficients $k = [k_1, \ldots, k_N]$. The solution w is a linear combination of eigenwalkers $e_1, \ldots, e_N$ with the obtained coefficients as the corresponding weights. Two calculation approaches are explored to find these coefficients.

3.2.1 Approach I. Since a 3D Fourier representation of the walker w can be represented as a linear combination of eigenwalkers with a set of coefficients k_n, substituting equation 3.3 into equation 3.5 gets

$$\hat{w} = C\left(\overline{w} + \sum_{n=1}^{N} k_n e_n\right). \tag{3.6}$$

Denoting $\hat{\bar{w}} = C\bar{w}$ and $\hat{e}_n = Ce_n$, equation 3.6 can be rewritten as

$$\hat{w} - \hat{\bar{w}} = \sum_{n=1}^{N} k_n \hat{e}_n, \tag{3.7}$$

or, using matrix notation, as

$$\hat{w} - \hat{\bar{w}} = \hat{E}k. \tag{3.8}$$

Equation 3.8 contains 150 equations with N unknown coefficients $k = [k_1, \ldots, k_N]$. N is smaller than R, the number of walkers. For our calculations, we used a value of $N = 30$. Equation 3.8 is therefore an overdetermined linear equation system. We approximate a solution according to a least-square criterion using the pseudo-inverse

$$k = (\hat{E}^T\hat{E})^{-1}\hat{E}^T(\hat{w} - \hat{\bar{w}}). \tag{3.9}$$

3.2.2 Approach II. In actual situations, due to measurement noise of 2D motion sequences and incompleteness of training examples, the obtained coefficients from least-squares minimization may cause the 3D reconstruction to be far beyond the range of the training data. In order to avoid this overfitting, Bayes' rule is used to make a trade-off between quality of match and prior probability.

According to Bayes' rule, a posterior probability is

$$p\left(k|\hat{w}\right) \propto p\left(k\right) p\left(\hat{w}|k\right). \tag{3.10}$$

The prior probability $p\left(k\right)$ reflects prior knowledge of the possible values of the coefficients. It can be calculated from the above linear motion model. Assuming a normal distribution along each of the eigenvectors, the prior probability is

$$p\left(k\right) \propto \prod_{i=1}^{N} \exp\left(-k_i^2/\left(2\lambda_i\right)\right) = \exp\left(-\sum_{i=1}^{N} k_i^2/\left(2\lambda_i\right)\right), \tag{3.11}$$

where $\lambda_i, i = 1, \ldots, N$ is the eigenvalue corresponding to the eigenwalker e_i. Because the variations along the eigenvectors are uncorrelated within the set of walkers, products can be used.

Our goal is to determine a set of coefficients $k = [k_1, \ldots, k_N]^T$ for the 2D Fourier representation $\hat{w}$ with maximum probability in the 3D linear space. We assume that each dimension of the 2D Fourier representation $\hat{w}$ is subject to uncorrelated gaussian noise with a variance σ^2. Then the likelihood of measuring $\hat{w}$ is given by

$$p(\hat{w}|k) \propto \exp(-\|\hat{w} - \hat{\bar{w}} - \hat{E}k\|^2/(2\sigma^2)), \tag{3.12}$$

where $\hat{\overline{w}} = C\overline{w}$ is the projection of average motion and $\hat{E} = CE$ denotes the projection of eigenwalkers. According to equation 3.10, the posterior probability is

$$p\,(k|\hat{w}) \propto \exp\left(-\left\|\hat{w} - \hat{\bar{w}} - \hat{E}k\right\|^2 / \left(2\sigma^2\right) - \sum_{i=1}^{N} k_i^2 / \left(2\lambda_i\right)\right). \tag{3.13}$$

The maximization of equation 3.13, corresponding to a maximum a posteriori (MAP) estimate, can be found by minimizing the following equation:

$$k_{MAP} = \arg\min_k \left(\left\|\hat{w} - \hat{\bar{w}} - \hat{E}k\right\|^2 / \left(2\sigma^2\right) + \sum_{i=1}^{N} k_i^2 / \left(2\lambda_i\right)\right). \tag{3.14}$$

The optimal estimate k_{opt} is then calculated (see appendix A for computational details),

$$k_{opt} = k_{MAP} = \left[diag\left(\frac{\sigma^2}{\lambda}\right) + \hat{E}^T\hat{E}\right]^{-1} \hat{E}^T\left(\hat{w} - \hat{\bar{w}}\right). \tag{3.15}$$

where $diag\,(x)$ is an operation to produce a square matrix, whose diagonal elements are the elements of the vector x. We can see when eigenvalue $\lambda(i)$ is large, the effect of the variance σ^2 on the coefficients $k(i)$ in the optimal estimate is less than those with small eigenvalues.

After the coefficient k is estimated using the two proposed approaches, the missing information $\tilde{w} = (\tilde{p}_0, \tilde{p}_1, \tilde{q}_1, \tilde{p}_2, \tilde{q}_2)$ of the Fourier representation of a 2D walker is synthesized in terms of the respective linear combination of 3D eigenwalkers $e_1, \ldots, e_N$. The reconstructed Fourier representation w is the combination of the 2D Fourier representation $\hat{w}$ and the reconstructed $\tilde{w}$:

$$w = ([\hat{w}, \tilde{w}]) = ([\hat{p}_0, \tilde{p}_0], [\hat{p}_1, \tilde{p}_1], [\hat{q}_1, \tilde{q}_1],\ [\hat{p}_2, \tilde{p}_2], [\hat{q}_2, \tilde{q}_2]). \tag{3.16}$$

3D periodic human motion is obtained from the reconstructed Fourier representation w by using equation 3.1.

4 Experiments

Using walking data acquired with a marker-based motion capture system, as described in section 2, we conducted two experiments. 2D test sequences were created by orthographic projection of the 3D walkers onto a vertical plane. We used viewpoints that ranged from the left profile view to the frontal view and from there to the right profile view in 1 degree steps. In the first experiment, we tried to reconstruct the missing, third dimension assuming that the view angle of the 2D test walker was known. In the second experiment, we assumed the horizontal viewpoint to be unknown and tested whether we could retrieve it together with the 3D motion. In both experiments, we used a complete leave-one-out cross-validation procedure: one walker is set aside for testing when creating the linear motion model and later projected onto it and evaluated. This is then repeated for every walker in the data set.

4.1 Model Construction. Fourier representations were created separately for each walker. On average, the fundamental frequency accounted for 91.9% of the total variance, and the second harmonic accounted for another 6.0%, which meant that the first two harmonics explained 97.9% of the overall postural variance of a walker. The sets of all the Fourier representations were now submitted to a PCA. The first 30 principal components accounted for 95.5% of the variance of all the Fourier representations, and they were chosen as eigenwalkers in the linear motion model (see Figure 2).

4.2 Motion Reconstruction. For approach II, in order to determine the MAP for a given 2D motion sequence, we calculated the optimal variance σ^2 (see equation 3.15). Since the 2D motion sequence was the orthographic projection of a 3D walker onto a vertical plane, the actual motion in the missing dimension was known and could be compared directly to the reconstructed data. We define an absolute error for each joint, comparing the reconstructed data with the original data, by

$$E_{abs}(j) = \frac{1}{T} \sum_{t=1}^{T} (p^j(t) - \tilde{p}^j(t))^2, \tag{4.1}$$

where $p^j(t)$ and $\tilde{p}^j(t)$ are the original and reconstructed data for the jth ($j = 1, 2, \ldots, 15$) joint in the missing dimension. The absolute reconstruction error of one walker is the average value of the absolute errors for the

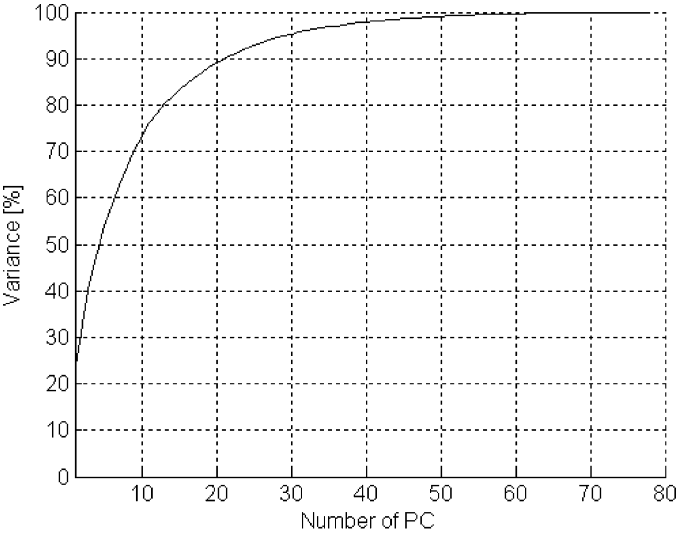

Figure 2: The cumulative variance covered by the principal components computed across 80 walkers.

15 joints:

$$E_a = \frac{1}{15} \sum_{j=1}^{15} E_{abs}(j). \tag{4.2}$$

We calculated the average reconstruction errors for the 80 walkers for variances (σ^2 in equation 3.12) ranging from 1 to 625 mm^2. Figure 3 illustrates the results. The optimal variance for all the walkers was about $\sigma^2 = 70$ mm^2. This value was used for approach II in all subsequent experiments.

To calculate a 3D Fourier representation, we projected the 2D Fourier representation onto the linear motion model and got a set of coefficients by the two proposed approaches. In order to qualitatively evaluate the reconstruction, we displayed all the subjects reconstructed from three viewing angles: 0, 30, and 90 degrees. For each demonstration, the original and reconstructed motion data were superimposed and displayed from a direction orthogonal to the direction along which motion data were missing. Figure 4 illustrates the original and the reconstructed motion sequences for one subject from the three viewing angles. This figure shows five equally spaced postures in one motion cycle. The original and the reconstructed walking sequences appear very similar in the corresponding postures, and there are no obvious differences between the two calculation approaches.

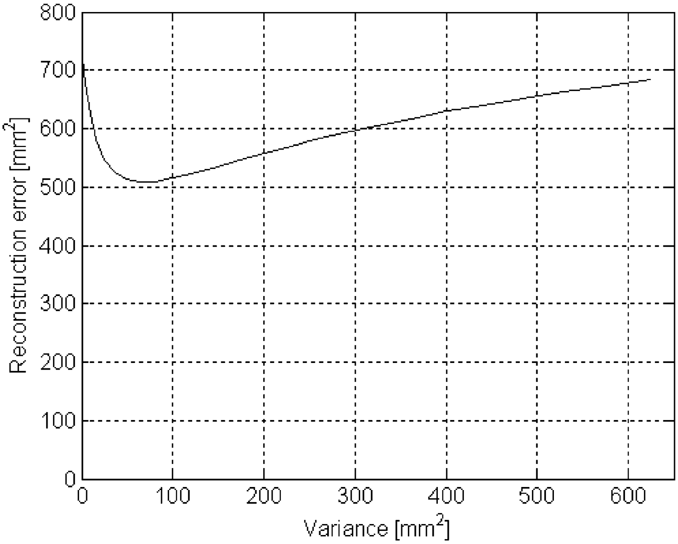

Figure 3: The effect of variance σ^2 on average reconstruction error.

The same results were inspected visually in all 240 demonstrations; no obvious differences could be seen.

In order to provide a more quantitative evaluation, equation 4.2 was used to calculate the absolute reconstruction error. From different viewpoints, the 2D projections have different variances, and therefore a relative reconstruction error is also defined for each joint,

$$E_{rel}(j) = \frac{1}{T\sigma^2}\sum_{t=1}^{T}(p^j(t) - \tilde{p}^j(t))^2, \tag{4.3}$$

where σ^2 is the average overall variance of the 15 joints in the missing dimension. The relative reconstruction error of one walker is the average value of the relative errors for the 15 joints:

$$E_r = \frac{1}{15}\sum_{j=1}^{15} E_{rel}(j). \tag{4.4}$$

The absolute and relative reconstruction errors were calculated for all walkers and for different viewpoints using the two approaches. 3D walking data were reconstructed at 1 degree intervals from the left profile view to the right profile view. Figure 5 shows absolute average errors, relative average errors, and overall variance of the missing dimension as a function

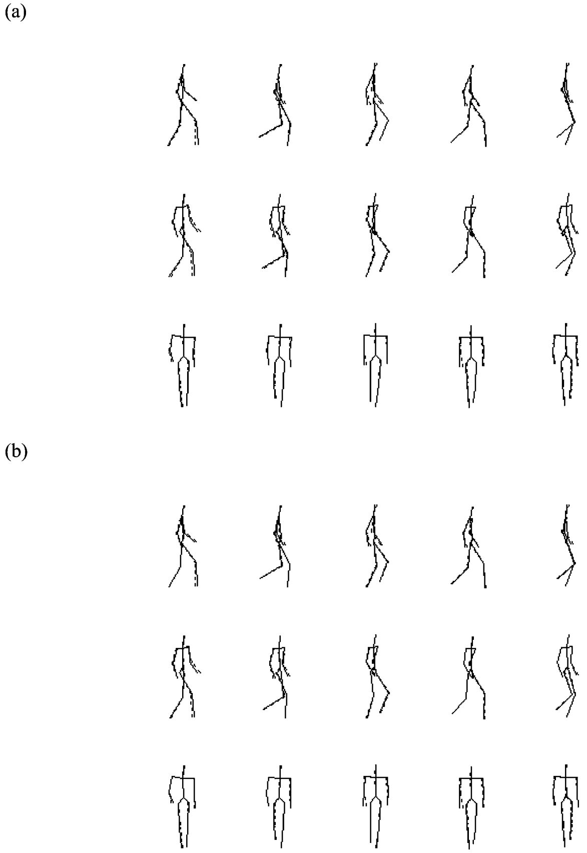

Figure 4: Original and reconstructed walking data from three viewing angles (0, 30, and 90 degrees, corresponding to the top, middle, and bottom rows, respectively). The dashed lines and the solid lines are the original and the reconstructed postures in the motion sequence, respectively. (a) Approach I. (b) Approach II.

of horizontal viewpoint (−90 to 90 degrees). The solid and dashed curves in Figures 5a and 5b correspond to approach I and approach II, respectively. It is obvious that when prior probability is involved in the calculation, the error of 3D motion reconstruction is about 30% smaller than the error by approach I. As can be seen from Figure 5a, there are different absolute reconstruction errors in the negative and positive oblique viewpoints, which implies that human walking is slightly asymmetric. The very small asymmetry of variances in Figure 5c also confirms this point.

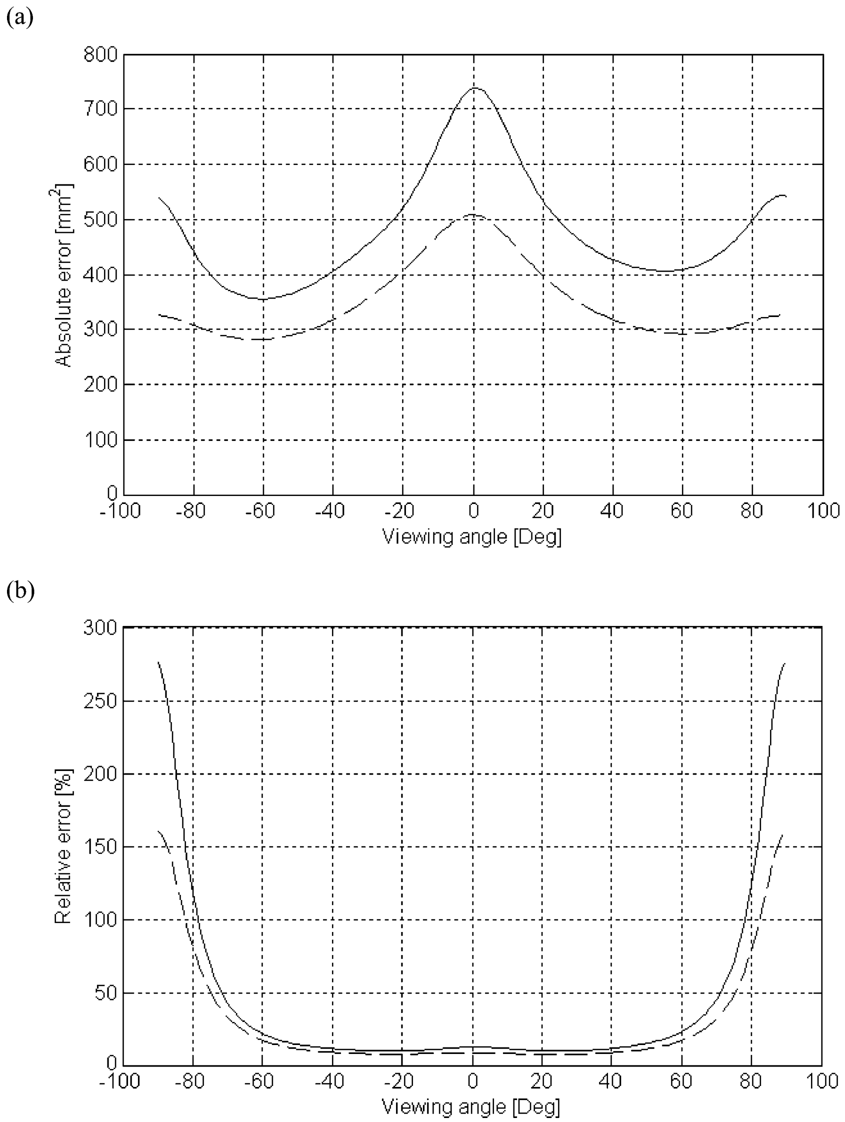

Figure 5: Illustration of the average reconstruction errors from different viewpoints of −90 to 90 degrees for 80 walkers. (a) Absolute errors by two approaches. (b) Relative errors by two approaches. (c) Variance.

(c)

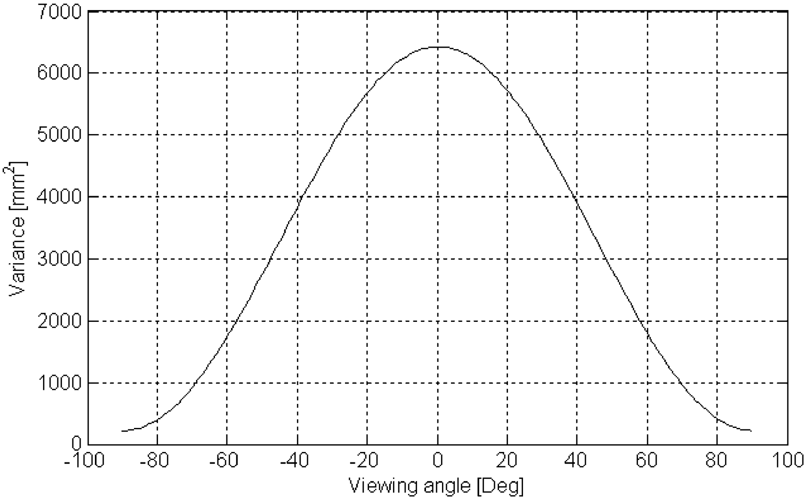

Figure 5: Continued

Figure 5b illustrates that reconstruction based on frontal and oblique viewpoints produces small relative reconstruction errors, while reconstruction from the profile view generates large relative reconstruction errors. The main reason for that is the fact that the overall variance that needs to be recovered from the profile view is relatively small anyway (see Figure 5c), and a reasonable absolute reconstruction error becomes a large relative error. The poor reconstruction in profile views and the good reconstruction in frontal views fit with data on human performance in biological motion recognition tasks (Troje, 2002a; Mather & Murdoch, 1994).

4.3 Viewpoint Reconstruction. While we assumed the projection matrix C (see equation 3.5) to be known in the first experiment, we include the horizontal viewpoint as an unknown variable, which is subject to reconstruction in the second experiment. Assuming that a 2D walking sequence has a constant walking direction and that the viewpoint is rotated only about the vertical axis (that is, it is a horizontal viewpoint), we can estimate the view angle using the rotated average posture $\hat{\bar{w}}(\alpha)$ and the rotated eigenwalkers $\hat{E}(\alpha)$ and then retrieve the missing motion data. Equation 3.8 and 3.14 can be represented as

$$(\alpha_{opt}, k_{opt}) = \underset{\alpha,k}{\arg\min} \left\| \hat{w} - \hat{\bar{w}}(\alpha) - \hat{E}(\alpha)k \right\|, \tag{4.5}$$

$$(\alpha_{opt}, k_{opt}) = \arg\min_{\alpha,k} \left(\left\| \hat{w} - \hat{w}(\alpha) - \hat{E}(\alpha)k \right\|^2 / \left(2\sigma^2\right) + \sum_{i=1}^{N} k_i^2 / (2\lambda_i) \right). \tag{4.6}$$

The optimum solution of this minimization problem over the coefficient k and view angle α can be found by solving a nonlinear overdetermined system. Details can be found in appendix B. The missing data are the linear combination of eigenwalkers using the optimal coefficients k_{opt}. As in the first experiment, a leave-one-out cross-validation procedure was applied to all the walkers. The average reconstructed angles from different viewpoints by the two approaches are illustrated in Figure 6. The experimental results show that we can precisely obtain the view angles from motion sequences with unknown viewpoints. Here, Bayesian inference does not provide an advantage. It is possible to identify and recognize human gait from 2D image sequences with different viewpoints by using the obtained view angle and missing data.

5 Conclusions and Future Work

We have investigated and evaluated the problem of reconstructing 3D periodic human motion from 2D motion sequences. A linear motion model, based on a linear combination of eigenwalkers, was constructed from Fourier representations of walking examples using PCA. The Fourier representation of a particular walker is a compact description of human walking, so not only does it allow us to find temporal correspondence by adjusting the frequency and phase, but it also reduces computational expenditure and storage requirements. Projecting 2D motion data onto this model could find a set of coefficients and view angles for test data with unknown viewpoints. Two calculation approaches were explored to determine the coefficients. One was based on a least-square minimization using the pseudoinverse; the other used prior probability of training examples to calculate a MAP by using Bayes′ rule. Experiments and evaluations were made on walking data. The results and quantified error analysis showed that the proposed method could reasonably reconstruct 3D human walking data from 2D motion sequences. We also verified that the linear motion model could estimate parameters like the view angle from 2D motion sequences with unknown constant walking direction. By applying Bayes′ rule to 3D motion reconstruction, we got better performance in reconstruction.

We developed and tested the proposed framework on a relatively simple data set: discrete marker trajectories obtained from motion capture data from walking human subjects. We also made a series of assumptions that simplified algorithmic and computational demands. In particular, we assumed that the projection matrix was either completely known or that only

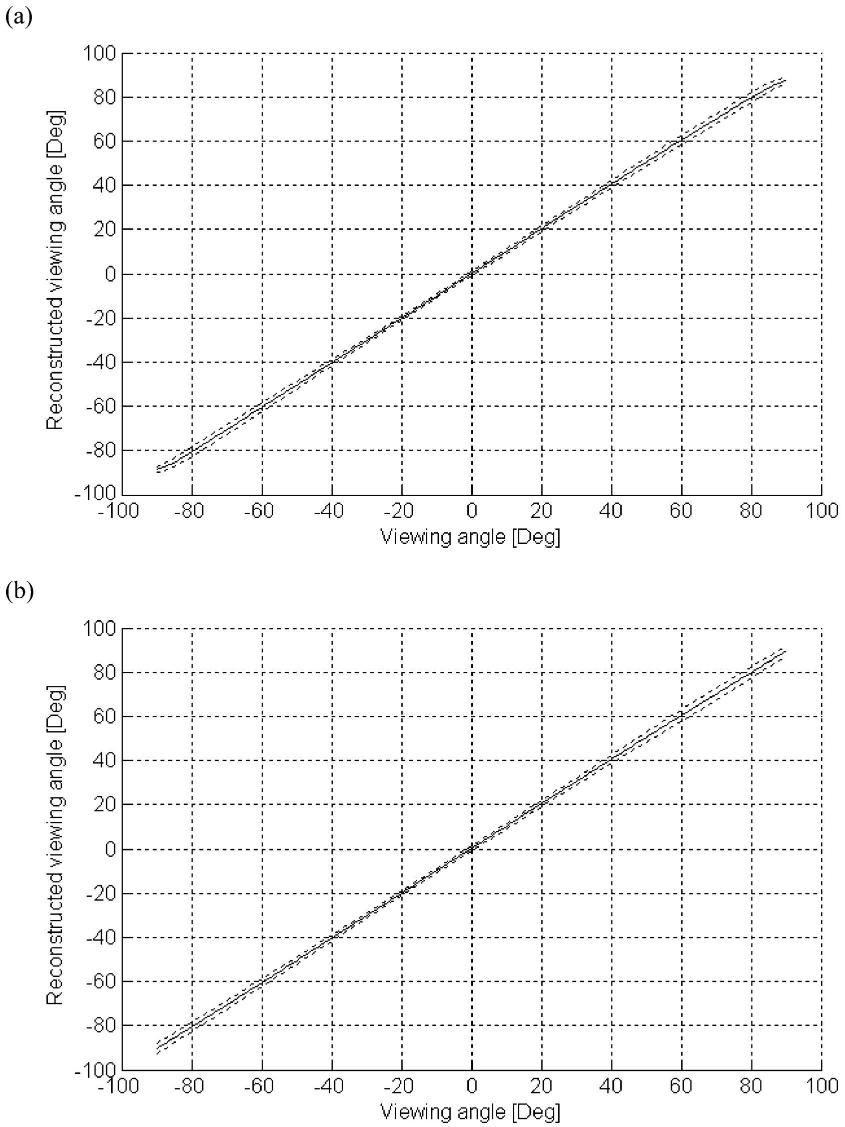

Figure 6: Illustration of the average reconstruction view angles from different viewpoints of −90 to 90 degrees for 80 walkers. The solid line is the average reconstructed angles, and the dashed lines are the average standard deviations. (a) Approach I. (b) Approach II.

a single parameter, the horizontal viewpoint, had to be recovered. Using the proposed framework to process real-world video data would require a number of additional steps and expansions to the model.

The computer vision problem of tracking joint locations in video footage remains challenging. Markerless tracking, however, has become a huge research field, and promising advances have been made along several lines. A discussion of the area is beyond the scope of this letter, but we are positive that solutions to the problem will be available in the near future.

Another constraint that could be relaxed in future versions of this work is the knowledge about the projection matrix. The fact that the single parameter that we looked at here, horizontal viewing angle, was recovered effortlessly and with high accuracy shows that the redundancy in the input data is still very high. Great numbers of unknown parameters would require more sophisticated optimization procedures, but we consider it very likely that the system would stay overdetermined, providing enough constraints to find a solution. Recently, Chai and Hodgins (2005) simultaneously extracted the root position and orientation from the vision data captured from two cameras.

There is another challenge when we switch from periodic actions like walking and running to nonperiodic movements. Establishing temporal correspondence across our data set was easy: mapping one sequence onto another simply required scaling and translation in time. Establishing correspondence between nonperiodic movements generally requires nonlinear time warping. A number of different methods are available to achieve this task, ranging from simple landmark registration to dynamic programming and the employment of hidden Markov models (see, e.g., Ramsay & Silverman, 1997), time-warping algorithm (Bruderlin & Williams, 1995), spatiotemporal correspondence (Giese & Poggio, 2000; Mezger, Ilg, & Giese, 2005), and registration curve (Kovar & Gleicher, 2003).

The main purpose of this work was to emphasize the richness of the redundancy inherent in motion data and suggest ways to employ this redundancy to recover missing data. In our example, the missing data were the third dimension, which gets lost when projecting the 3D motion data onto a 2D image plane. The same approach, however, could be used to recover markers that become occluded by other parts of the body or even to generate new marker trajectories at locations where a real marker cannot be placed (for instance, inside the body).

Appendix A: Solving an Overdetermined System for Coefficients

In order to calculate the optimal estimates k, we assume the following equation:

$$E = \left(\hat{w} - \hat{\hat{w}} - \hat{E}k\right)^2 / \left(2\sigma^2\right) + \sum_{i=1}^{N} k_i^2 / (2\lambda_i). \tag{A.1}$$

It can be written as

$$E = \left(\left(\hat{w} - \hat{\overline{w}} \right)^2 - 2 \left\langle k, \hat{E}^T \left(\hat{w} - \hat{\overline{w}} \right) \right\rangle + \left\langle k, \hat{E}^T \hat{E} k \right\rangle \right) / \left(2\sigma^2 \right) + \sum_{i=1}^{N} k_i^2 / \left(2\lambda_i \right). \tag{A.2}$$

According to the optimum,

$$0 = \nabla E = \left(-2\hat{E}^T \left(\hat{w} - \hat{\overline{w}} \right) + 2\hat{E}^T \hat{E} k \right) / \left(2\sigma^2 \right) + \sum_{i=1}^{N} 2k_i / \left(2\lambda_i \right), \tag{A.3}$$

so

$$\hat{E}^T \hat{E} k + \sigma^2 \sum_{i=1}^{N} k_i / \lambda_i = \hat{E}^T \left(\hat{w} - \hat{\overline{w}} \right). \tag{A.4}$$

The solution is

$$k = \left[diag\left(\frac{\sigma^2}{\lambda} \right) + \hat{E}^T \hat{E} \right]^{-1} \hat{E}^T \left(\hat{w} - \hat{\overline{w}} \right). \tag{A.5}$$

Appendix B: Solving an Overdetermined System for Coefficients and Angle

For solving equation 3.5, we define an objective function, whose return value is a 150-dimensional vector. Every element of this vector is

$$diff(i) = \hat{w}(i) - \hat{\overline{w}}(i)(\alpha) - \hat{E}(i)(\alpha) \cdot k, \text{for } i = 1, \ldots, 150 \tag{B.1}$$

where $\hat{E}(i)(\alpha)$ is a row vector with the number of used eigenwalkers. The first 75 elements in this vector can be calculated by the following equation:

$$diff(i) = \hat{w}(i) - \cos\alpha \times \left(\overline{w}_x + E_x \cdot k \right) - \sin\alpha \times \left(\overline{w}_y + E_y \cdot k \right), \tag{B.2}$$

where $i = 1, \ldots, 75$, $\overline{w}_x$, E_x, and $\overline{w}_y$, E_y are the corresponding x- and y-coordinates of average walkers and eigenwalkers, respectively. The values of the remaining 75 elements can be calculated by

$$diff(i) = \hat{w}(i) - \overline{w}_z - E_z \cdot k, \tag{B.3}$$

where $i = 76, \ldots, 150$, and $\overline{w}_z$ and E_z are the corresponding z-coordinates of average walkers and eigenwalkers, respectively.

Finally, the obtained 150-dimensional vector is sent to the subroutine lsqnonlin in Matlab 6.5. The "large-scale: trust-region reflective Newton" optimization algorithm was used in this subroutine. The returned values are the optimum coefficients and the viewing angle.

For equation 4.6, we calculate the coefficients for the viewing angle from 0 to 180 degrees by using equation 3.15 and then substitute them into equation 4.6. We can determine the optimum coefficients and view angle by comparing the 181 obtained values.

Acknowledgments

We gratefully acknowledge the generous financial support of the German Volkswagen Foundation and the Canada Foundation for Innovation. We thank Tobias Otto for his help in collecting the motion data and Daniel Saunders for many discussions and his tremendous help in the final editing of the manuscript. The helpful comments from the anonymous reviewers are gratefully acknowledged; they greatly helped us improve this letter.

References

Aggarwal, A., & Triggs, B. (2004). Learning to track 3D human motion from silhouettes. In *Proceedings of the 21st International Conference on Machine Learning*. Madison, WI: Omni Press.

Aggarwal, J. (2003). Problems, ongoing research and future directions in motion research. *Machine Vision and Applications, 14*, 199–201.

Aggarwal, J., & Cai, Q. (1999). Human motion analysis: A review. *Computer Vision and Image Understanding, 73*, 428–440.

Arikan, O., & Forsyth, D. (2002). Interactive motion generation from examples. *ACM Transactions on Graphics, 21*(3), 483–490.

Blanz, V., & Vetter, T. (2003). Face recognition based on fitting a 3D morphable model. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 25*, 1063–1074.

Bowden, R. (2000). *Learning statistical models of human motion*. Paper presented at the IEEE Workshop on Human Modeling Analysis and Synthesis, CVPR 2000, South Carolina.

Bowden, R., Mitchell, T., & Sarhadi, M. (2000). Non-linear statistical models for the 3D reconstruction of human pose and motion from monocular image sequences. *Image and Vision Computing, 18*, 729–737.

Bruderlin, A., & Williams, L. (1995). Motion signal processing. *Proceedings of SIGGRAPH, 95*, 97–104.

Buxton, H. (2003). Learning and understanding dynamics scene activity: A review. *Image and Vision Computing, 21*, 125–136.

Chai, J., & Hodgins, J. (2005). Performance animation from low-dimensional control signals. *ACM Transactions on Graphics, 24*(3), 686–696.

Cunado, D., Nixon, M., & Carter, J. (2003). Automatic extraction and description of human gait models for recognition purposes. *Computer Vision and Image Understanding, 90*, 1–41.

Dariush, B. (2003). Human motion analysis for biomechanics and biomedicine. *Machine Vision and Applications, 14*, 202–205.

Gavrila, D. (1999). The visual analysis of human motion: A survey. *Computer Vision and Image Understanding, 73*, 82–98.

Giese, M., & Poggio, T. (2000). Morphable models for the analysis and synthesis of complex motion patterns. *International Journal of Computer Vision, 38*, 59–73.

Grochow, K., Martin, S., Hertzmann, A., & Popović, Z (2004). Style-based inverse kinematics. *ACM Transactions on Graphics, 23*(3), 522–531.

Howe, N., Leventon, M., & Freeman, W. (1999). Bayesian reconstruction of 3D human motion from single-camera video. In S. Solla, T. Leen, & K.-R. Müller (Eds.), *Advances in neural information processing systems, 12* (pp. 820–826). Cambridge, MA: MIT Press.

Jepson, A., Fleet, D., & El-Maraghi, T. (2003). Robust online appearance models for visual tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 25*, 1296–1311.

Kakadiaris, I., & Metaxas, D. (2000). Model-based estimation of 3D human motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*, 1453–1459.

Karaulova, I., Hall, P., & Marshall, A. (2002). Tracking people in three dimensions using a hierarchical model of dynamics. *Image and Vision Computing, 20*, 691–700.

Kovar, L., & Gleicher, M. (2003). Flexible automatic motion blending with registration curves. In *Proceedings of Symposium on Computer Animation* (pp. 214–224). Aire-la-Switzerland: Eurographics Association.

Leventon, M., & Freeman, W. (1998). *Bayesian estimation of 3D human motion from an image sequence*. (Tech. Rep. No. 98-06). Cambridge, MA: Mitsubishi Electric Research Laboratory.

Mather, G., & Murdoch, L. (1994). Gender discrimination in biological motion displays based on dynamic cues. *Proceedings of the Royal Society of London, Series B: Biological Sciences, 258*, 273–279.

Mezger, J., Ilg, W., & Giese, M. (2005). Trajectory synthesis by hierarchical spatiotemporal correspondence: Comparison of different methods. In *Proceedings of the ACM Symposium on Applied Perception in Graphics and Visualization* (pp. 25–32). New York: ACM Press.

Moeslund, T., & Granum, E. (2001). A survey of computer vision-based human motion capture. *Computer Vision and Image Understanding, 81*, 231–268.

Ning, H. Z., Tan, T. N., Wang, L., & Hu, W. M. (2004). Kinematics-based tracking of human walking in monocular video sequences. *Image and Vision Computing, 22*, 429–441.

Ong, E., & Gong, S. (2002). The dynamics of linear combinations: Tracking 3D skeletons of human subjects. *Image and Vision Computing, 20*, 397–414.

Ramsay, J., & Silverman, B. (1997). *Functional data analysis*. New York: Springer.

Ren, L., Shakhnarovich, R., Hodgins, J., Pfister, H., & Viola, P. (2005). Learning silhouette features for control of human motion. *ACM Transactions of Graphics, 24*, 1303–1331.

Rosales, R., Siddiqui, M., Alon, J., & Sclaroff, S. (2001). Estimating 3D body pose using uncalibrated cameras. In *Proceedings of the Computer Vision and Pattern Recognition Conference* (Vol. 1, pp. 821–827). Los Alamitos, CA: IEEE Computer Society Press.

Sminchisescu, C., & Triggs, B. (2003). Kinematic jump processes for monocular 3D human tracking. In *Proceedings of the Computer Vision and Pattern Recognition Conference* (Vol. 1, pp. 69–76). Los Alamitos, CA: IEEE Computer Society Press.

Troje, N. F. (2002a). Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of Vision, 2*, 371–387.

Troje, N. F. (2002b). The little difference: Fourier based gender classification from biological motion. In R. P. Würtz & M. Lappe (Eds.), *Dynamic perception* (pp. 115–120). Berlin: Aka Press.

Unuma, M., Anjyo, K., & Takeuchi, R. (1995). Fourier principles for emotion-based human figure animation. *Proceedings of SIGGRAPH, 95*, 91–96.

Urtasun, R., & Fua, P. (2004a). 3D human body tracking using deterministic temporal motion models. In *Proceedings of the Eighth European Conference on Computer Vision*. Berlin: Springer-Verlag.

Urtasun, R., & Fua, P. (2004b). 3D tracking for gait characterization and recognition. In *Proceedings of the Sixth International Conference on Automatic Face and Gesture Recognition*. (pp. 17–22). Los Alamitos, CA: IEEE Computer Society Press.

Wang, L., Hu, W. M., & Tan, T. N. (2003). Recent developments in human motion analysis. *Pattern Recognition, 36*, 585–601.

Wang, L., Tan, T. N., Ning, H. Z., & Hu, W. M. (2003). Silhouette analysis-based gait recognition for human identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 25*, 1505–1518.

Yacoob, Y., & Black, M. (1999). Parameterized modeling and recognition of activities. *Computer Vision and Image Understanding, 73*, 232–247.

Received April 25, 2005; accepted August 16, 2006.