Available online at www.sciencedirect.com
SCIENCE DIRECT
Image and Vision Computing 23 (2005) 1264–1276
image AND vision COMPUTING
www.elsevier.com/locate/imavis

Representing cyclic human motion using functional analysis☆

Dirk Ormoneit [a], Michael J. Black [b,*], Trevor Hastie [c], Hedvig Kjellström [d]

[a] Marshall Wace LLP, 1/11 John Adam Street, London WC2N 6HT, UK
[b] Department of Computer Science, Brown University, 115 Waterman Street, Box 1910, Providence, RI 02912, USA
[c] Department of Statistics, Stanford University, Stanford, CA 94305, USA
[d] Department of IR Systems, Swedish Defence Research Agency, SE-164 90 Stockholm, Sweden

Received 9 February 2004; received in revised form 5 August 2005; accepted 6 September 2005

Abstract

We present a robust automatic method for modeling cyclic 3D human motion such as walking using motion-capture data. The pose of the body is represented by a time-series of joint angles which are automatically segmented into a sequence of motion cycles. The mean and the principal components of these cycles are computed using a new algorithm that enforces smooth transitions between the cycles by operating in the Fourier domain. Key to this method is its ability to automatically deal with noise and missing data. A learned walking model is then exploited for Bayesian tracking of 3D human motion.

Keywords: Human motion; Functional data analysis; Missing data; Singular value decomposition; Principal component analysis; Motion capture; Tracking

1. Introduction

Increasing amounts of three-dimensional human motion data are available from commercial motion-capture systems and are used for a variety of applications. For example, they provide an input for the animation of virtual characters in computer graphics (see [21] for an overview), for biometric person identification [8,17,55,56], or for the analysis of gait abnormalities in medicine [38,62]. The focus of much of this work has been on cyclic human motions such as walking and running. A common approach for modeling such motions uses principal component analysis (PCA) to represent them in a low-dimensional linear subspace that captures the natural variations among people and activities [5,9,26,30,41, 53,56,65,66]. Our work addresses some of the difficulties in such an approach by providing a fully automatic method for learning PCA models of human motion. In particular, we show how the cycle length can be automatically determined and provide a new algorithm for learning PCA models with missing data that is tailored to *cyclic* data. The resulting set of techniques can be exploited by any motion modeling approach that first represents cyclic human motion using PCA. To illustrate the methods we learn a probabilistic model for human walking and we apply it to the problem of human motion recovery from a sequence of 2D video images.

In many applications, the human body is approximated by a collection of articulated limbs (Fig. 1) that form a kinematic tree. The motion of the body can then be thought of as a collection of time-series describing the joint angles as they evolve over time. A key difficulty for the modeling of these body angles is that each time-series has to be decomposed, or segmented, into sequences of meaningful actions (or 'movemes' [12]) prior to statistical analysis. For example, in the case of repetitive human motion such as walking, motion sequences decompose naturally into a sequence of similar motion cycles. The exact nature of this decomposition is unknown a priori and needs to be estimated from the motion data. In this work, we present a new set of tools that carry out this identification automatically. These tools also allow us to automatically compute the mean and the principal components of the individual cycles. While PCA models of human gait are, by now, quite standard [5,9,30,41,53,56,65,66] a number of technical difficulties have not previously been addressed. One

☆Part of this work has been presented at the joint IMS/Beroulli Conference in Guanajuato, Mexico, May 2000 and Neural Information Processing Systems 2000.

* Corresponding author Tel.: +1 401 863 7637.
E-mail addresses: d.ormoneit@mwam.com (D. Ormoneit), black@cs.brown.edu (M.J. Black), hastie@stat.stanford.edu (T. Hastie), hedvig@nada.kth.se (H. Kjellström).
URLs: http://www.cs.brown.edu/people/black/ (M.J. Black), http://www-stat.stanford.edu/~trevor/ (T. Hastie), http://www.nada.kth.se/~hedvig/ (H. Kjellström).

0262-8856/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.imavis.2005.09.004

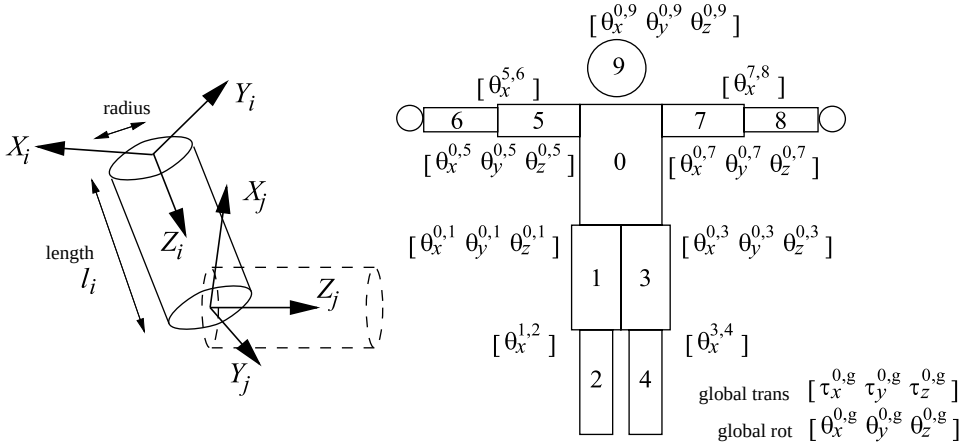

Fig. 1. The human body is represented as a kinematic tree consisting of articulated, tapered, cylinders with 25 degrees of freedom (DOF); six degrees for the translation and rotation of the torso and 19 relative joint angles expressed here as Euler angles. Each limb, i, has a local coordinate system with the Z_i axis directed along the limb. Joints have up to 3 angular DOF, expressed as relative rotations ($\theta_x^{j,i}$, $\theta_y^{j,i}$, $\theta_z^{j,i}$) between body parts i and j.

important issue involves dealing with missing information in the motion time-series due to occlusion. Missing data is common in even high-quality motion-capture data and commercial systems rely on ad hoc, semi-automated methods to fill in what is missing. Such methods may introduce bias or remove fine temporal structure from the data. Additionally, in the case of cyclic activities, one needs to enforce smooth transitions between cycles. To deal with these problems, we develop a new iterative method for functional principal component analysis that is based on a singular value decomposition (SVD) in the Fourier domain. The result is a fully automatic method that takes a database of cyclic human motions and produces a statistical model that is suitable for various applications. This automated method ignores missing data and consequently avoids problems of bias introduced by post hoc data manipulation.

To demonstrate the application of such a model we explore its use in recovering 3D human motion from 2D image sequences. In this tracking application, the principal components serve to define a low-dimensional representation of the human 3D poses in a state-space model that treats the 2D video images as observations. We specify the transition probabilities (or motion prior) for this state-space model using the singular values of the learned motion model and we also specify an image likelihood term based on a generative model of the image appearance. This generative model is non-linear and the posterior probability density over human poses is non-Gaussian. Consequently, we exploit approximate Bayesian inference techniques for temporal filtering. Specifically, we apply a particle filter [23,29] where the posterior distribution is represented by a discrete set of samples and is propagated over time using Monte Carlo techniques. The resulting algorithm is able to track human subjects in monocular video sequences and to recover their 3D motion under changes in their pose and against complex unknown backgrounds.

The remainder of this paper is organized as follows: In Section 2 we briefly review previous work on human motion modeling and tracking. In Section 3, we present details of the data set under investigation and we describe the sequence alignment procedures that divide the motion-capture data into cycles. In Section 4, we explain our algorithm for functional principal component analysis and we illustrate the method with human walking motions. The application for tracking is described briefly in Section 5 to illustrate the ideas; for a detailed description, the reader is referred to [41,52,53].

2. Previous work

There has been substantial previous work on the modeling of human motion data in the statistical and computer science communities. Statistical representations of time-series data using functional analysis are described in detail in [45]. An interesting treatment of the statistical analysis of human motion data in a medical context can be found in [40] while Faraway [20] applies functional analysis to a problem in ergonomics. For a good review of human motion modeling see [56].

Much of the work on gait analysis has focused on 2D image representations such as silhouettes, raw images, or features derived directly from images [6,7,16,18,27,31,37,39]. Here, we focus on 3D body joint angles, which have also received a great deal of attention for biometric person identification [8,17,27,55] emotion modeling [2,56] and tracking [30,53,59,60,65].

A common approach to model repetitive 3D human motion data is to hand-segment and align the data in contrast to the automated alignment procedures suggested here. Previous approaches for automated alignment can be found, for example, in [46]. Based on this segmentation the individual walking ‘cycles’ can be modeled using various statistical techniques, e.g. those described in [4] or [47].

We focus here on modeling 3D joint angles for tracking, rather than recognition of, human motion. The simplest such models place constraints on the smooth change in joint angles [25,61]. These models are derived from biometric studies of human motion [13,48,49] or learned from 3D motion-capture data. The learned statistical properties may be captured by wavelets [43], principal component analysis (PCA) [5,9,26,30,53,56,65,66], polynomial basis functions [22],

Fourier components [17,56,58], dynamical models [42,63], vector quantization [26,36], phase-space constraints [14], or various types of hidden Markov models (HMM) [9–11,36].

Cyclic human motion, in particular, has received special attention [1,35,50,56,65] where the focus is often on recognition of individuals based on 'signatures' in their gait or recognition of deviations from the norm (e.g. as a result of carrying a heavy object). For cyclic articulated motion, Pullen and Bregler [43] use a frequency decomposition of joint angles with a learned, non-parametric, kernel density estimate of the conditional statistics across frequency bands. Sampling from this model produces synthetic repetitive motions with natural variation.

In previous work on principal component analysis of human motion data, the 3D motion curves corresponding to particular activities were typically hand-segmented and aligned [6,53,65]. In contrast, this paper details an automated method for segmenting the data into individual activities, aligning activities from different examples, modeling the statistical variation in the data, dealing with missing data, enforcing smooth transitions between cycles, and deriving a probabilistic model suitable for a Bayesian interpretation. PCA with missing data has been dealt with in other contexts [33,51] but not for the case of cyclic motions.

Murase and Sakai [37] use PCA to reduce the dimensionality of silhouette images of walkers and then represent activities as trajectories of eigenspace coefficients. This differs from the approach of representing trajectories themselves in a low-dimensional subspace. Similarly, Huang et al. [27] build an eigenspace of spatio-temporal templates and exploit this for gait recognition. The approach does not deal explicitly with 3D joint angle data or the problem of detecting and modeling individual motion cycles.

In contrast to PCA approaches, Cunado et al. [17] extract 2D motion information about the upper leg from video and then model the upper leg motion using a Fourier series. The Fourier basis is used in contrast to the learned PCA basis for specific activities. An advantage of the PCA approach is that it automatically captures covariation between limbs when the model is trained using multiple joints.

Troje [56] takes a slightly different approach by performing PCA on marker positions (not 3D joint angles) to derive a low-dimensional posture model. The time-series of postures in this low-dimensional space are then fit by sinusoids. To capture variability across subjects and behaviors Troje then performs PCA on the reduced-dimension time-series. In contrast to a learned PCA representation of the time-series, the sinusoid representation may remove important subtly from the motion. Our approach goes beyond this to deal with the automated modeling of cyclic motions. We work directly with times series and do not make the sinusoidal assumption. Key to our approach is the ability to automatically deal with imperfect and missing data. Finally we apply the model to 3D human tracking.

More recently, there has been an emphasis on non-parametric models of human motion that essentially assemble motion clips sampled from a large database of motions [3,32,34,44,64]. Like the analogous texture synthesis techniques, these methods provide detailed synthesis but do not generalize well to motions not present in the database. The functional analysis approach here provides a probabilistic model that can generalize within a particular class of motions and hence may be more appropriate for tracking. Parameterized methods also have advantages for gait analysis and recognition.

Bayesian methods for tracking 3D human motion have been used previously in [15,19,26,42,53,54]. In these methods, temporal curves learned with PCA can be thought of as providing a prior probability distribution over valid human motions. Given the high dimensionality of the human body, this prior constrains the possible motions to lie on a far lower dimensional manifold. This makes the problem of pose estimation with Monte Carlo sampling techniques more tractable. Further illustrating the value of PCA-based models of human motion, recent work has explored using these models and deterministic optimization methods to recover 3D motion from video [59,60].

3. Human motion data

In this work, we develop a modeling procedure for periodic motion sequences. By definition, periodic motion is composed of repetitive cycles which constitute a natural unit of statistical modeling and which must be identified in the training data prior to building a model. For example, Fig. 2 illustrates one particular cycle of a walking sequence. Frequently, this segmentation is carried out manually in an error-prone and burdensome procedure (see, for example, [45,65]). In this

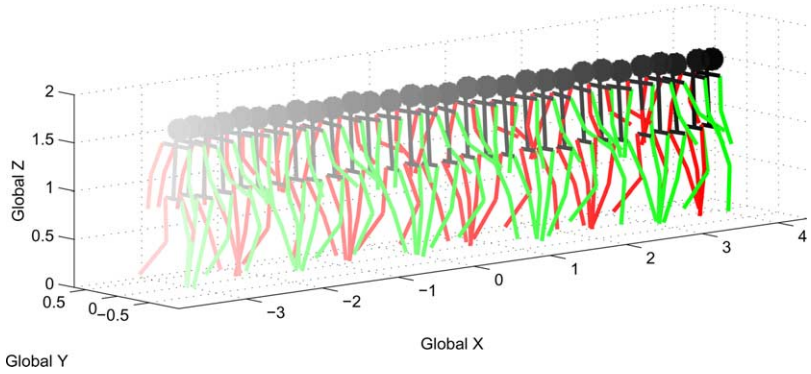

Fig. 2. Example walking sequence synthesized from motion-capture data (torso motion and relative joint angles).

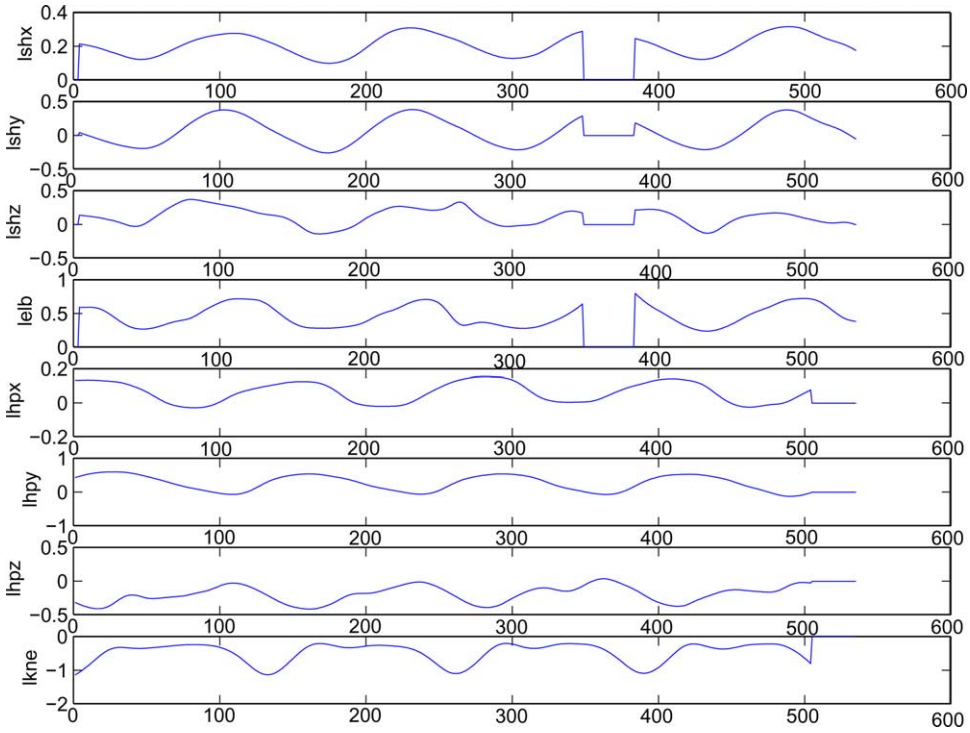

Fig. 3. Time-series representation of the angles in the walking sequence. We show the left shoulder (lshx, lshy, lshz), elbow (lelb), hip (lhpx, lhpy, lhpz), and knee angles (lkne). Note regions of missing data (e.g. between approximately time instants 300–400 and after 500).

section, we present alignment algorithms that segment the data automatically. Based on this alignment, the mean and the principal components of the cycle data are computed as a statistical model.

3.1. Motion-capture data

Three-dimensional human motion data were collected using a commercial Vicon motion-capture system. Four subjects (professional dancers) performed a variety of activities while wearing retro-reflective markers. Commercial software was used to reconstruct the 3D trajectories of the markers and the relative joint angles of the limbs over time; see Fig. 2 for an example of one walking sequence. Eight walking sequences (two per subject) were recorded at a sampling frequency of 120 frames/second and the lengths of the sequences range from about 500 to 5000 frames. It is important to emphasize that the modeling techniques suggested in this paper can be applied to a much wider range of repetitive motions.

For the purpose of our analysis, we focus on the 19 relative joint angles describing the limb motion over time (refer to Fig. 1). In doing so we eliminate any global position information though this could be modeled as well. In Fig. 3 we show a subset of eight of these 19 time-series that characterize the left body-half.

Formally, we let T_i denote the length of the ith motion sequence and we use $t=1,\ldots,T_i$ as a time index. Similarly, $m=19$ is the number of angles in each motion sequence and $a=1,\ldots,m$ indicates a particular angle. The ith motion sequence can be written as[1]

$$Z_i(t) \equiv \{z_{i,a}(t) | a = 1, \ldots, m\} \quad \text{for } t = 0, \ldots, T_i.$$

Associated with each sequence we have the indicator set

$$I_{i,a} \equiv \{t \in \{1, \ldots, T_i\} | z_{i,a}(t) \text{ is not missing}\}.$$

That is, $I_{i,a}$ is a set of time indices that labels non-missing pieces of information in $z_{i,a}$ and we have one such set for each sequence i and for each angle a. Missing observations arise frequently in our data set because some markers may be occluded by other body parts during portions of a motion. This is very common, even in high-quality commercial data. The capturing system reports an angle of zero for some of the position coordinates in this case. Typically occlusions last for several frames, which prevents the imputation of interpolated values using neighboring observations.

Next, we describe a procedure to estimate alignment parameters that segment periodic motion sequences into cycles. There are several ways to approach this issue in principle (see, for example [46]). For the data described in Section 3.1, however, it seems particularly relevant that the procedure handles missing observations and that it produces reliable estimates even for the relatively short human motion sequences we have at our disposal, consisting of only 3–4 cycles per sequence. The procedure consists of two steps. For

[1] The data Z_i is not to be confused with the reference coordinate Z_j for limb j in Fig. 1.

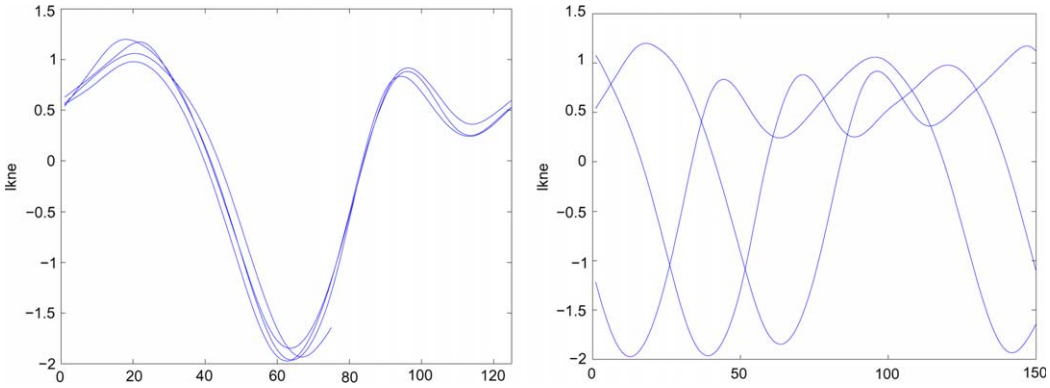

Fig. 4. Wrapped walking cycles for the left knee angle. Left: a near-optimal cycle length of $p=125$ results in a strong signal to noise ratio ($\text{signal}_{\text{lkne}}=254.3$). Right: an incorrect cycle length results in no clear alignment; e.g. $p=150$ ($\text{signal}_{\text{lkne}}=0.4527$).

each motion sequence, we first estimate its cycle length, p, using a wrapping procedure; second, we estimate an offset parameter, o, which describes the relative shift of that sequence using an iterative algorithm. In practice, these parameters vary due to different speeds and different starting positions of the individuals in the capture sequence. Based on p and o, the individual motion sequences can be transformed into a common reference domain for further analysis.

3.2. Estimation of the cycle length

To illustrate the effect of choosing the correct cycle length p consider the 'wrapping' procedure shown in Fig. 4 which shows the left knee angle during several walking cycles. Each candidate value p defines a segmentation of the original time-series into a sequence of segments that can be represented in the common domain $\{1,\ldots,p\}$. Formally, let the 'projection index' associated with p be defined according to $\xi_p(t) \equiv t \bmod p$, so that ξ_p 'folds' the original sequence into the domain $\{1,\ldots,p\}$. For each $k \in \{1,\ldots,p\}$ we will consider all those elements of Z_i that are non-missing and that are projected onto k. The indices of these elements are collected in the sets $I_{i,a}(k) \equiv \{j \in I_{i,a} | k = \xi_p(j)\}$ for $k=1,\ldots,p$. Also, let $|I_{i,a}(k)|$ be the cardinality of $I_{i,a}(k)$. Then the mean of the observations mapped onto k and the overall mean of sequence i can be written as

$$\bar{z}_{i,a}(k) \equiv \frac{1}{|I_{i,a}(k)|} \sum_{j \in I_{i,a}(k)} z_{i,a}(j), \qquad \bar{Z}_i \equiv \frac{1}{|I_{i,a}|} \sum_{j \in I_{i,a}} z_{i,a}(j).$$

The following magnitudes measure the signal- and the noise-content of the projected sequence:

$$\text{noise}_{i,a}(p) \equiv \frac{\sum_{j \in I_{i,a}} (z_{i,a}(j) - \bar{z}_{i,a}(\xi_p(j)))^2}{|I_{i,a}| - p} \tag{1}$$

$$\text{signal}_{i,a}(p) \equiv \frac{\sum_{k=1}^{p} (\bar{z}_{i,a}(k) - \bar{Z}_i)^2}{(p-1)p/|I_{i,a}|} \tag{2}$$

Specifically, $\text{noise}_{i,a}$ is an estimate of the average square deviation of the aligned cycles from the mean cycle, $\bar{z}_{i,a}$; it can be interpreted as the variation in the data that is not explained by $\bar{z}_{i,a}$. On the other hand, $\text{signal}_{i,a}$ is an estimate of the mean variation of the mean cycle, $\bar{z}_{i,a}$, with respect to the overall mean, $\bar{Z}_i$, and measures the signal intensity. Both $\text{noise}_{i,a}$ and $\text{signal}_{i,a}$ are normalized so as to account properly for the degrees of freedom of the estimates and to produce estimates whose expectation is independent of p. Because it is natural to favor values of p that produce simultaneously high signal- and low noise-contributions, we combine (1) and (2) into the 'signal-to-noise ratio'[2]:

$$\text{snr}_i(p) \equiv \sum_a \frac{\text{signal}_{i,a}(p)}{\text{noise}_{i,a}(p)}. \tag{3}$$

In our algorithm we try candidate values for p in a suitable range and choose the maximum with respect to (3) as our estimate of the cycle-length. Note that (3) comprises the accumulated contributions from all angles in our data set. In Fig. 5 we show the individual signal-to-noise ratios for a subset of the angles as well as the accumulated signal-to-noise ratio (3) as functions of p. Note the sharp peak of these values around the optimal cycle length $p=126$. The peak around the optimal value is more expressed in some sequences than in others. While the variance in the angles differs for each joint, the maximal peak remains the same. For example, the left Z hip angle (lhpz) varies much less than the hip Y angle and the resulting signal-to-noise ratio for the Y-axis is eight times that of the Z-axis. Note also that the signal-to-noise ratio of the artificially generated white noise series in the first row is approximately constant, warranting the unbiasedness of our approach with respect to changing values of p.

3.3. Estimation of the offset parameter

The folding procedure described above computes an estimate of the optimal cycle length $p(i)$ for each sequence and stores these values in an array of length n. In our second step, we use this array to align multiple sequences in a common domain by rescaling each individual sequence i according to

[2] Alternative estimation approaches based on spectral analysis cannot easily be applied due to the missing data. Even if there were no missing observations, the statistical efficiency of these estimates might be low due to the relatively small number of frames per period. That is, the conditions of the Sampling Theorem may not be fulfilled so as to guarantee a sufficiently small approximation error.

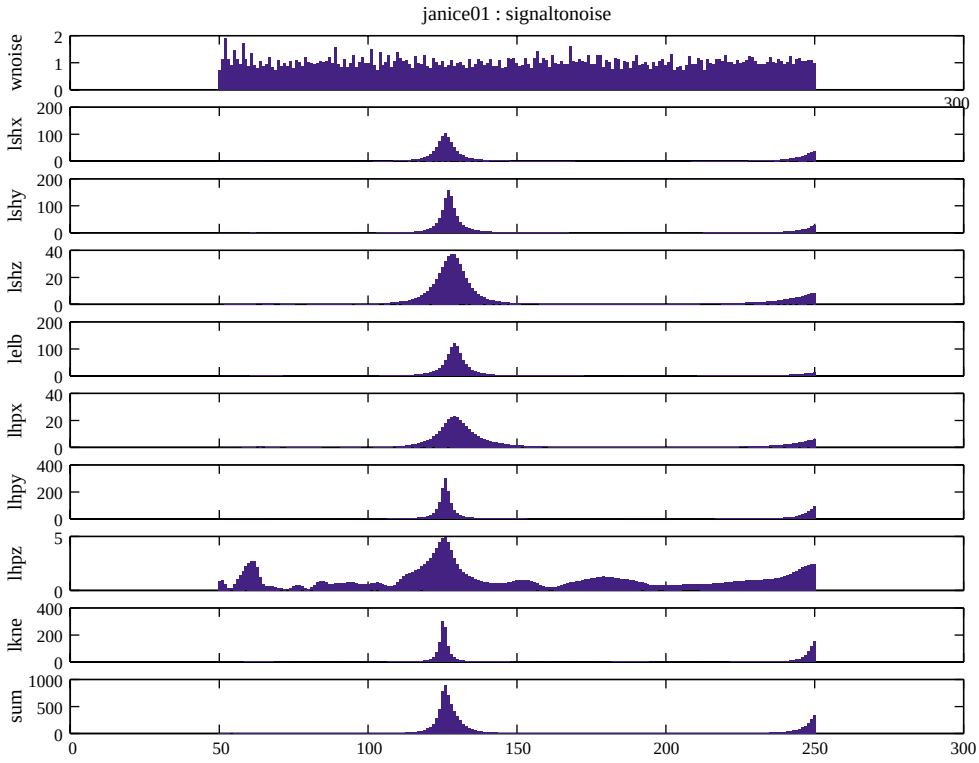

Fig. 5. Signal-to-noise ratio of a representative set of angles as a function of the candidate period length. We show the same angles as in Fig. 3. The top series corresponds to a white noise signal. The bottom series shows the accumulated signal-to-noise ratio (3). The minimum overall variation was obtained for $p=126$, ((snr = 8.986505e + 02)).

$p(i)$ and by shifting it according to an offset parameter $o(i)$. More specifically, for each i we define the following mapping of the time-series index $t=0,\ldots,T_i$ into the domain [0, 1]:

$$\zeta_i(t) \equiv \frac{(t-o(i)) \bmod p(i)}{p(i)}.$$

Then we construct offset estimates $o(1), o(2),\ldots,o(n)$ so as to allow for the best approximation of the time-series in terms of a common *reference model*, $r(\tau)$, where $\tau \in [0, 1]$. In other words, for each i we require that the time-series $z_{i,a}(t)$ deviates as little as possible from its reference value, $r(\zeta_i(t))$. We need a separate reference model $r_a(\tau)$ for each angle a and each of these models belongs to a pre-specified class of functions, $\mathcal{R}$. As a measure of deviation between $z_{i,a}(t)$ and $r(\zeta_i(t))$ we use their squared distance.

There are two technical issues complicating this strategy. First, because the motion sequences are typically asynchronous after rescaling using the estimated cycle-lengths, we have to adapt our definitions of the reference signal and of the noise level to a continuous-time framework. Second, we are facing a computational problem. An exhaustive search for the optimal offset-combination requires $O\left(\prod_{i=1}^{n} p(i)\right)$ evaluations of the signal-to-noise criterion, which is clearly infeasible in practice. Instead, we suggest the iterative procedure illustrated in Fig. 6 to compute an approximate solution. We initialize the offset values to zero in Step 1, and we define the reference signal r_a in Step 2 so as to minimize the deviation with respect to the aligned data. Note that missing data are discarded for estimation. Next, we choose the offsets of all sequences so that they minimize the prediction error with respect to the reference signal (Step 3). By contrast to the exhaustive search, this operation requires only $O\left(\sum_{i=1}^{n} p(i)\right)$ comparisons. Because the solution of the first iteration may well be suboptimal, we construct an improved reference signal using the current offset estimates, and use this signal in turn to improve the offset estimates. Repeating these steps, we obtain an iterative optimization algorithm that is terminated if the improvement falls below a given threshold (Step 4). Since Steps 2 and 3 both decrease the prediction error, it is clear that the algorithm converges.

Fig. 7 shows eight sequences of a walking motion, aligned using this procedure. As a function class for the reference signal,

1. Initialize offset values. For each motion sequence $i = 1,\ldots,n$, let $o(i) := 0$.
2. From a given function class $\mathcal{R}$, choose the minimum least-squares fit with respect to the aligned data. For $a = 1,\ldots,m$:
$$r_a := \arg\min_{r\in\mathcal{R}} \sum_{i=1}^{n} \sum_{j\in I_{i,a}} \left[z_{i,a}(j) - r\left(\zeta_i(j)\right)\right]^2.$$
3. Update the offset parameters. For $i = 1,\ldots,n$:
$$o(i) := \arg\min_{o} \sum_{a} \sum_{j\in I_{i,a}} \left[z_{i,a}(j) - r_a\left(\zeta_i(j)\right)\right]^2.$$
4. Stop, if the performance improvement is below 10^{-6}. Otherwise, goto Step 2.

Fig. 6. Iterative algorithm for the computation of the optimal offset parameters.

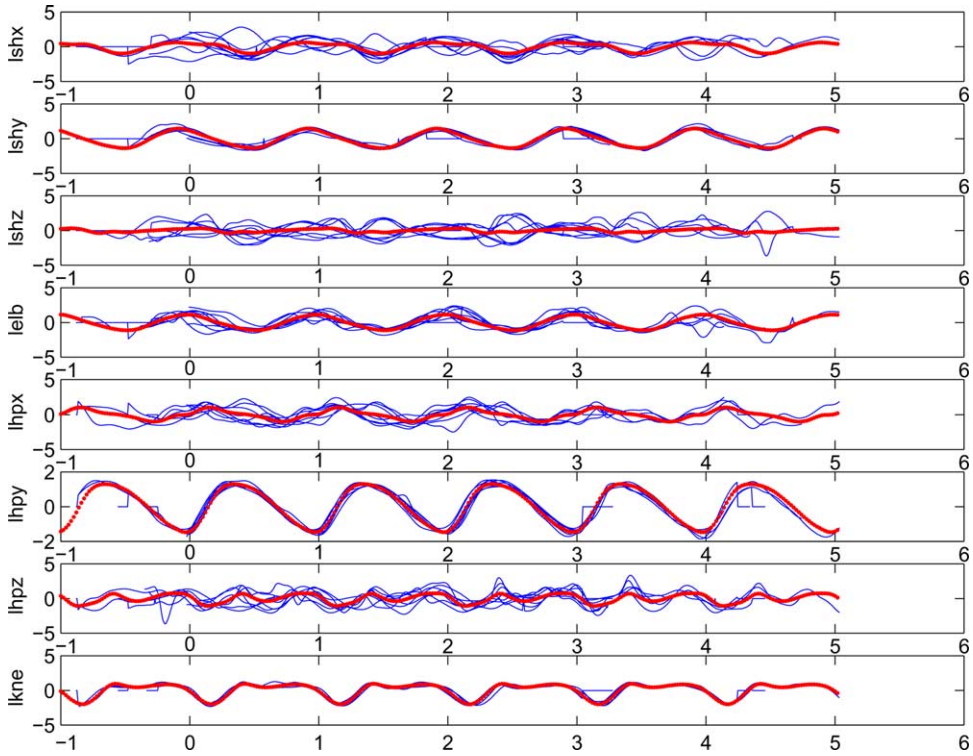

Fig. 7. Aligned representation of eight walking sequences (blue). The red curve denotes repetitions of the periodic regression spline estimate (convergence after five steps).

$\mathcal{R}$, we chose periodically constrained regression splines in our implementation, i.e. the zeroth, first, and second derivatives of the spline are designed so as to coincide at the boundaries (for details, see Appendix A). Without these constraints, the concatenated reference signal in Fig. 7 would be discontinuous at the transitions between cycles.

3.4. Transformation into reference domain

Next, we use this aligned representation to transform the data into a new reference domain. This step is necessary because even though the data are aligned in Fig. 7, they are still sampled asynchronously due to the different alignment parameters of each sequence. The goal of the synchronization is to represent the data in a matrix of fixed size that will be the basis of our principal component algorithm in Section 4. Note also that using the algorithm of Section 3.3 we effectively estimated a continuous-time model of the mean curve (the red curve) in Fig. 7. For synchronization we will estimate continuous-time models for each cycle of the individual time-series (the blue curves) and then transform this model into a new reference domain. We will denote the transformed cycle data as 'motion segments' and we will use the reference domain $\{0,\ldots,\mathcal{T}\}$. The sequence $z_{i,a}$ contains $K_i = \lceil (T_i - o(i))/p(i) \rceil + 1$ cycles for $o(i) \neq 0$ which are labeled $\tilde{z}_{k,a}$. Here, $k = 1,\ldots,K$ ranges over the combined segments from all sequences $i = 1,\ldots,n$ and $K = \sum_{i=1}^{n} K_i$. (We also count the partially filled 'ends' on the left and on the right, which are filled up with missing data labels.) $\tilde{z}_{k,a}$ and $\tilde{I}_{k,a}$ denote the kth segment and the corresponding indicator set, respectively, i.e. $\tilde{I}_{k,a}$ contains the indices of the non-missing elements of $\tilde{z}_{k,a}$. As a function estimate, we compute a (unconstrained) regression spline estimate in the domain [0, 1] for each $k = 1,\ldots,K$ and for each angle:

$$f_{k,a} := \arg\min_{f \in \mathcal{S}} \sum_{j \in \tilde{I}_{k,a}} \left[\tilde{z}_{k,a}(j) - f\left(\frac{j}{p(i)}\right) \right]^2.$$

Here, $\mathcal{S}$ is the class of functions that can be expressed as linear expansions of the basis vectors $\phi_1,\ldots,\phi_K$ defined in Appendix A. This gives a representation of the segments in a continuous domain. Next, we define the reference domain $\{0, \ldots, \mathcal{T}\}$ and we transform the data into the reference domain according to $f_{k,a}$:

$$\bar{z}^{(1)}_{k,a}(j) := f_{k,a}\left(\frac{j}{\mathcal{T}}\right) \quad \text{for } j = 0, 1, \ldots, \mathcal{T}.$$

The superscript (1) is to distinguish the original data from various approximations in intermediate computing steps described below. For the principal component analysis, it is convenient to rearrange the data in $\bar{z}^{(1)}_{k,a}$ in matrix format. We define the following design matrix for this purpose:

$$X^{(1)} = \begin{pmatrix} \bar{z}^{(1)}_{1,1}(0) & \ldots & \bar{z}^{(1)}_{1,1}(\mathcal{T}) & \ldots & \bar{z}^{(1)}_{1,m}(0) & \ldots & \bar{z}^{(1)}_{1,m}(\mathcal{T}) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \bar{z}^{(1)}_{K,1}(0) & \ldots & \bar{z}^{(1)}_{K,1}(\mathcal{T}) & \ldots & \bar{z}^{(1)}_{K,m}(0) & \ldots & \bar{z}^{(1)}_{K,m}(\mathcal{T}) \end{pmatrix}.$$

That is, each row of $X^{(1)}$ is a concatenation of the segments corresponding to the angles $a=1,\ldots,m$ and the rows of $X^{(1)}$ correspond to the index $k=1,\ldots,K$. Below we use the symbols $\bar{z}_{k,a}^{(N)}$ and $X^{(N)}$ to refer to the same data depending on whether we prefer to describe a processing step using vector or matrix format.

4. Principal component analysis

Next, we derive a statistical model of the repetitive cycles in Fig. 7. For example, in 3D-animation, it may be desirable to model a mean motion and systematic deviations from the mean to assign personalities to virtual characters. Similarly, in motion tracking we wish to identify motion sequences that deviate from the mean and for activity recognition we must represent the unique characteristics of a person's motion and the natural variability.

As a more comprehensive description of the underlying probability distribution, we use the principal components of the synchronized time-series data. Here, an important question is whether to compute the principal components *jointly* over all the angles or *individually* for each angle. Clearly this decision depends on the amount of dependence between the individual angles in a particular type of motion. For repetitive sequences—in particular for walking—the individual angles are typically highly dependent, suggesting the joint modeling approach.

A new algorithm to estimate the principal components is shown in Fig. 8. The algorithm addresses the problems of previous approaches, which do not account for missing data or enforce continuity across cyclic movemes. In our approach, the data segments $\bar{z}_{k,a}^{(1)}$ defined above undergo a sequence of processing stages. We use the notation $\bar{z}_{k,a}^{(N)}$, $N=1,\ldots,6$ to denote the intermediate results of this stage-wise computation, and we interpret $\bar{z}_{k,a}^{(1)}$ as the first element of this sequence. Similarly, $X^{(N)}$ is the design matrix at stage N, which is composed of the data in $\bar{z}_{k,a}^{(N)}$, $k=1,\ldots,K$, $a=1,\ldots,m$. Our algorithm addresses several difficulties: first, because of the missing data in $X^{(1)}$, we cannot simply use a standard singular value decomposition (SVD) to obtain the principal components. As an alternative approach, we use an iterative approximation scheme suggested recently by [24,57] in the context of analyzing Gene Expression Arrays. Specifically, we alternate between an SVD step (2–5) and a data imputation step (Step 6), where each update is designed so as to decrease the matrix distance between the original data and their reconstruction. As an additional complication, we cannot compute the SVD directly because the principal components obtained in this manner could be non-periodic or even discontinuous. A pragmatic approach is to project the data onto a smooth and periodic basis prior to carrying out the SVD (Step 2). We choose a Fourier basis for this purpose, and we truncate the high-frequency components by keeping only the 20 leading coefficients. This gives a new coefficient matrix $X^{(3)}$. Next, we compute the optimal rank-q approximation to $X^{(3)}$ using a complex SVD in Steps 3 and 4, and we reconstruct the signal in the original domain in Step 5. In Step 6 we use the reconstructed values as improved estimates for the missing data in $X^{(2)}$, and then we repeat Steps 2–5 using these improved estimates. This iterative process is continued until the performance improvement falls below a given threshold.

Theorem 1. The Functional Principal Component Algorithm in Fig. 8 monotonically minimizes the matrix distance between the normalized design matrix and its reconstruction, $\|X^{(2)}-X^{(5)}\|$.

Proof. It is clear that the data imputation Step 6 decreases the reconstruction error because the differences between the imputations of missing values are set to zero. Steps 2–5 project $X^{(2)}$ onto the linear subspace spanned by the low-frequency Fourier basis. To see this, note that the Steps 2 and 5 can be

1. Compute the row-mean μ of $X^{(1)}$, and let $X^{(2)} := X^{(1)} - 1^T\mu$. Here 1 is a vector of ones. Initialize missing values in $X^{(2)}$ using zeros.
2. For each of the segments $\bar{z}_{k,a}^{(2)}$ in $X^{(2)}$, compute the first 20 Fourier coefficients and store the result in a new matrix, $X^{(3)}$.
3. Compute the Singular Value Decomposition of $X^{(3)}$:
$$X^{(3)} = USV^T.$$
4. Reconstruct the design matrix, using the rank q approximation to S:
$$X^{(4)} = US^qV^T. \quad (4)$$
5. Apply the Inverse Fourier Transform to obtain $X^{(5)}$.
6. Evaluate the reconstruction error $||X^{(2)} - X^{(5)}||$. Update the imputation of the missing values:
$$\bar{z}_{k,a}^{(2)}(j) = \bar{z}_{k,a}^{(5)}(j) \text{ for all } j \in \bar{I}_{k,a}.$$
7. If the error reduction is above 10^{-6} goto Step 2.
8. Return $X^{(6)} = X^{(5)} + 1^T\mu$.

Fig. 8. Functional PCA algorithm with data imputation and enforcement of smoothness constraints.

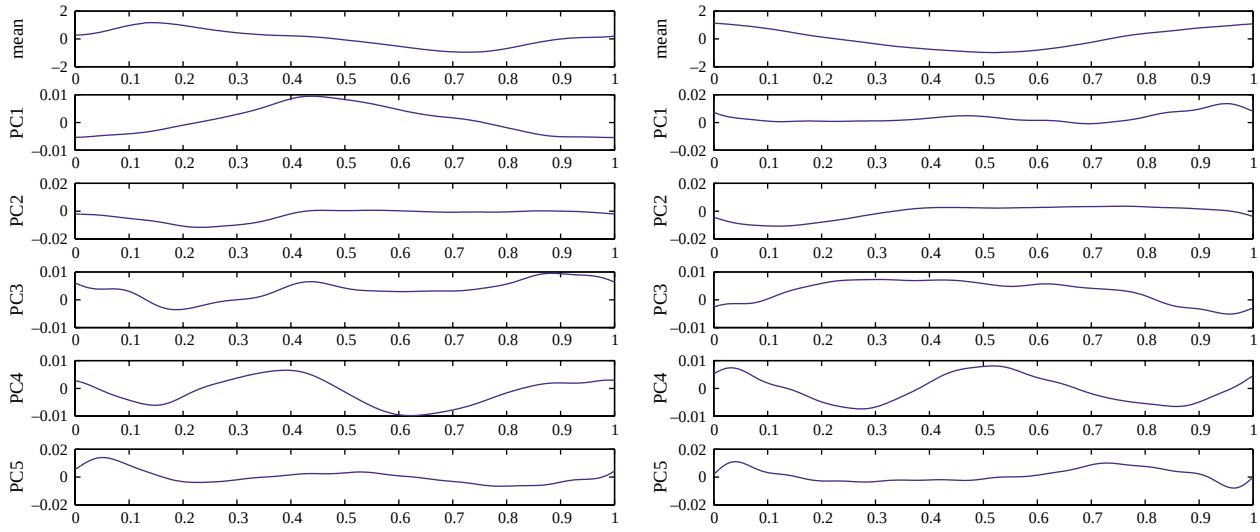

Fig. 9. Left: Mean and the four leading principal components of the left hip-angle during walking, estimated from eight sequences. The singular values corresponding to these principal components are 892.0, 410.5, 362.2, 290.6, 255.3. Right: The same data for the left elbow. The singular values of these data are 631.8, 513.3, 430.0, 356.8, 246.8.

written alternatively as matrix multiplications

$$X^{(3)} = X^{(2)}B \quad \text{and} \quad X^{(5)} = X^{(4)}B^{\mathrm{T}} = US^qV^{\mathrm{T}}B^{\mathrm{T}}.$$

Here, $\tilde{B} = (B, B_\perp)$ describes the angle-wise Fourier transform and B is the submatrix of $\tilde{B}$ that corresponds to the first 20 coefficients. $B_\perp$ determines the remaining coefficients and is orthogonal to B, i.e. $B^{\mathrm{T}}B_\perp = 0$. Because the Fourier basis is orthonormal, B and $B_\perp$ satisfy the conditions $B^{\mathrm{T}}B = I$ and $B_\perp^{\mathrm{T}}B_\perp = I$, respectively. The reconstruction error can be decomposed as follows:

$$\begin{aligned} ||X^{(5)} - X^{(2)}||^2 &= ||(X^{(5)} - X^{(3)})B||^2 + ||(X^{(5)} - X^{(2)})B_\perp||^2 \\ &= ||US^qV^TB^TB - X^{(2)}B||^2 + ||US^qV^TB^TB_\perp \\ &\quad - X^{(2)}B_\perp||^2 = ||US^qV^T - X^{(2)}||^2 + ||X^{(2)}B_\perp||^2. \end{aligned} \tag{5}$$

The first term in (5) is the rank-q reconstruction error of $X^{(2)}$, which is minimized by the values of U, S, and V computed in Step (4). Hence $X^{(5)}$ is the optimal rank-q approximation of $X^{(2)}$ within the class of matrices whose row-space is spanned by the rows of B. In other words, $X^{(5)}$ is a orthogonal projection of $X^{(2)}$ with respect to $\|\cdot\|$ and the second term in (5), which is independent of U, S, and V, corresponds to the residuals of this projection. Because $X^{(5)}$ minimizes $\|X^{(5)} - X^{(2)}\|$, it is superior to the $X^{(5)}$ from the previous iteration in particular. Hence both the data imputation and the SVD steps reduce $\|X^{(5)} - X^{(2)}\|$ which gives the desired result. □

In Fig. 9 we show the mean and the four principal components for various body angles estimated from the walking data. Because a common mean was computed for different individuals in this figure, the principal components can be interpreted as the main sources of variation *between* different individuals. In Fig. 10 we show the same principal components using different means for different individuals. Hence, the principal components characterize variations *within* different realizations by one and the same individual in this case. An intuitive way to interpret these data is by using animation. Here we add a relatively large multiple of each component to the mean cycle and visualize the resulting motion sequence. Corresponding movies are available on the World Wide Web[3]; they can be interpreted in terms of specific characteristics of some individuals in the experiment. Another interesting experiment is to artificially generate new motion sequences by using different random combinations of the principal components. Sequences generated in this manner are also available on our web site.

5. Application: motion tracking in video sequences

In this section, we briefly overview an application of the principal component model for the tracking of human motion in monocular video sequences. This example illustrates how the learned statistical model of human motion can be exploited to capture the prior probability of human motion for specific activities. Since our primary focus in this work is not on tracking but on the modeling aspects of human motion, we shall not discuss the details here; instead, we refer the reader interested in details to [52,53].

In our approach to tracking, the goal is to estimate the joint angles of the body and its 3D pose given a sequence of 2D image measurements, $\mathbf{J}_t \equiv \{J_t, \ldots, J_1\}$. Of course, 2D image data is insufficient in general to infer 3D information exactly. As an approximation, we interpret the human 3D poses as unobserved variables, ϕ_t, in a state-space model and we treat the image data J_t as observations. Here the principal components serve to define a low-dimensional representation of the human 3D poses, which is particularly important to reduce the computational burden and to incorporate prior information. Such

[3] http://www.cs.brown.edu/people/black/3Dtracking.html.

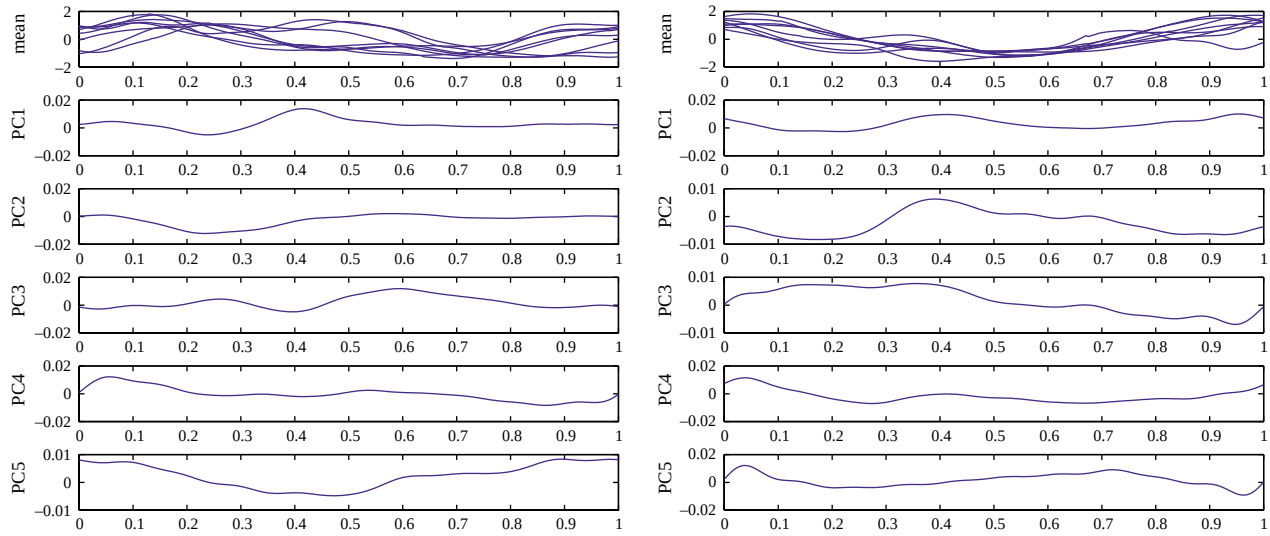

Fig. 10. Left: Mean and the four leading principal components of the left hip-angle during walking, estimated from eight sequences. The singular values corresponding to these principal components are 497.5, 465.9, 365.1, 273.9, 211.8. Right: The same data for the left elbow. The singular values of these data are 698.7, 542.5, 390.4, 285.4, 206.2.

additional information regarding the type of motion that can be expected in a video sequence is crucial to simplify the complex tracking task.

Formally, let $\theta(t) \equiv (\theta_a(t)|a = 1, \ldots, m)$ be a random vector summarizing the relative joint angles at time t; in other words, the motion sequence, $Z_i(t)$, at time t is now interpreted as the ith realization of $\theta(t)$. Under the modeling assumptions of the SVD in Fig. 8, $\theta(t)$ can be written in the form

$$\theta(t) = \tilde{\mu}(\psi_t) + \sum_{k=1}^{q} c_{t,k} v_k(\psi_t),$$

where v_k is the Fourier inverse of the kth column of V in (4), rearranged as a $\mathcal{T} \times m-$matrix; similarly, $\tilde{\mu}$ denotes the rearranged mean vector μ in Fig. 8. $v_k(\psi)$ is the ψth column of v_k, and the $c_{t,k}$ are time-varying coefficients. $\psi_t \in \{0, \mathcal{T}-1\}$ maps absolute time onto relative cycle positions or phases, and ρ_t denotes the speed of the motion such that $\psi_{t+1} = (\psi_t + \rho_t) \bmod \mathcal{T}$. Also, let $\mathbf{c}_t = (c_{t,1}, \ldots, c_{t,q})$ denote a vector of the q linear coefficients and let $\boldsymbol{\tau}_t^g$ and $\boldsymbol{\theta}_t^g$ represent the global 3D translation and rotation of the torso. Then body positions are characterized by the state-vector

$$\phi_t = (\mathbf{c}_t, \psi_t, \rho_t, \boldsymbol{\tau}_t^g, \boldsymbol{\theta}_t^g)^T.$$

The dynamics of a state-space model are described by the densities:

$p(\phi_t|\phi_{t-1})$ "Transition Model"
$p(J_t|\phi_t)$ "Observation Model".

In other words, the transition model, or transition density, $p(\phi_t|\phi_{t-1})$ characterizes the random change in ϕ_t from time $t-1$ to t, and the observation model, $p(J_t|\phi_t)$, characterizes the generation of images. Details of these densities are described in [41,52,53]. In tracking, interest at time t focuses on the conditional distribution $p(\phi_t|\mathbf{J}_t)$. This is because we would like to recover the probability distribution over the 3D body poses given the sequence of previous images. One possibility is to compute $p(\phi_t|\mathbf{J}_t)$ recursively using the *prediction* and *filtering* equations:

$$p(\phi_t|\mathbf{J}_{t-1}) = \int p(\phi_t|\phi_{t-1}) p(\phi_{t-1}|\mathbf{J}_{t-1}) \mathrm{d}\phi_{t-1} \tag{6}$$

$$p(\phi_t|\mathbf{J}_t) \propto p(J_t|\phi_t) p(\phi_t|\mathbf{J}_{t-1}). \tag{7}$$

(6) and (7) can be evaluated in closed form provided the transition and the observation models are linear with Gaussian noise. Unfortunately, the observation equation resulting from the generative image model is highly non-linear so that we have to resort to approximate inference for filtering. We use a particle filter for this purpose where $p(J_t|\phi_t)$ is represented as a weighted set of particles, or samples, where each sample represents the pose of the body in terms of the parameters of the learned body model. Tracking is achieved by propagating these particles in time (e.g. [23,28]).

To illustrate the method we show an example of tracking a walking person in a cluttered scene. On an Ultra 1 Sparc station the C++ implementation ran at a rate of approximately 1 frame/min. To visualize the posterior distribution we display the projection of the 3D model corresponding to the expected value of the model parameters. All parameters were initialized with a Gaussian prior at time $t = 0$.

Fig. 11 shows the tracking results for frames 0–50 of a walking sequence. Note that the legs of the model are better aligned with the image data than the arms. This is probably due to the fact that the arms are more often occluded by the torso, and thus more prior driven than the legs. Note also that the subject was not present in the training data. The small training set may lack sufficient variation to capture the motion of the test subject.

In parts of the cycle where large occlusion occurs (frame 30) the model has little image information, and starts to drift off the person. However, it recovers when a larger part of the body

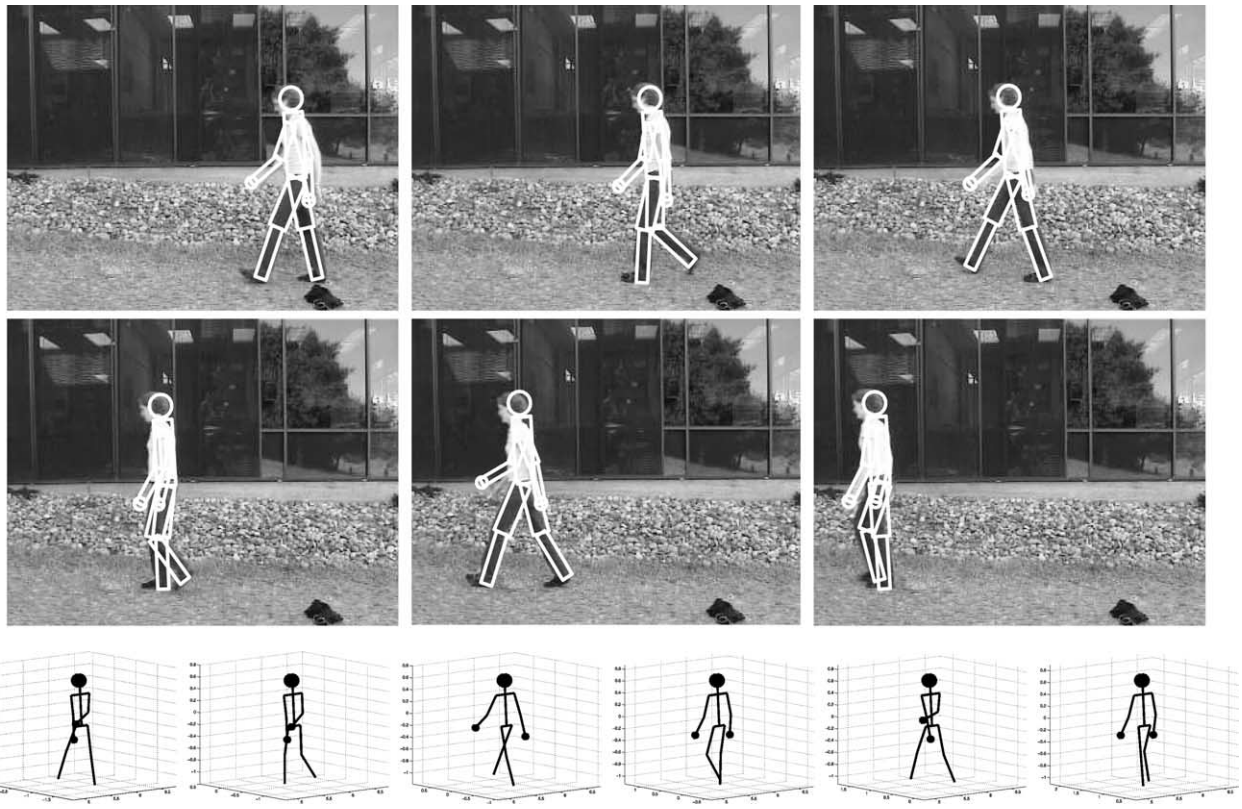

Fig. 11. Tracking of person walking, based on a particle filter with 10,000 samples. The two upper rows show frames 0, 10, 20, 30, 40, 50 in the sequence with the projection of the expected model configuration overlaid. The lower row shows the expected 3D configuration in the same frames.

is visible (frame 40). Video animations of these results can be found online at http://www.cs.brown.edu/people/black/3Dtracking.html.

For the interested reader, it is worth contrasting our stochastic Bayesian estimation method with a recent deterministic approach, which uses the same types of PCA models for tracking but exploits deterministic optimization methods [59,60].

6. Conclusion

Fields as diverse as graphics, gait recognition, and rehabilitation medicine require representations of cyclic human motions for synthesis or analysis and many of these methods rely on hand-crafted PCA representations. The approach described here provides an automated method for learning these periodic human motions from training data. Statistical methods are presented for detecting the length of the periods in the data, segmenting it into cycles, and optimally aligning the cycles. We also presented a novel principal component analysis technique for modeling the motion curves. The method copes with missing data and enforces smoothness between the beginning and ending of a motion cycle. The principal components serve to define a state-space model for the tracking of human motion in video sequences. We have demonstrated results for tracking a person in a cluttered image sequence but the method is applicable to any synthesis or analysis problem involving cyclic motions.

Acknowledgements

The work of Dirk Ormoneit was partly supported by a grant of the Deutsche Forschungsgemeinschaft (DFG). We are grateful to Michael Gleicher for generously providing the 3D motion-capture data used in our experiments. We thank Manolis Kamvysselis and David Fleet for many interesting discussions about human motion and visual tracking. David Fleet also made significant contributions to the 3D human tracking algorithm.

Appendix A. Periodic regression splines

We consider a regression spline estimate in the domain [0, 1] with fixed knots $\check{x}_k \in \{0.1, 0.2, \ldots, 0.9\}$. The regression spline, $f(x)$, is a linear expansion of the basis functions $\phi_k(x)$, $k=0,\ldots,K$ (for details, see Section 3.2.5 in [45]):

$$\phi_0(x) \equiv 1, \quad \phi_1(x) \equiv x, \quad \phi_2(x) \equiv x^2, \quad \phi_3(x) \equiv x^3,$$

$$\phi_{k+3}(x) \equiv (x - \check{x}_k)_+^3.$$

These basis functions are summarized in the function $\Phi(x) \equiv (\phi_1(x),\ldots,\phi_K(x))^{\mathrm{T}}$ so that $f(x) = \Phi(x)^{\mathrm{T}}\beta$ using a $K+1$ vector of coefficients β. Our objective is to enforce boundary constraints of the form $f^{(i)}(0) = f^{(i)}(1)$ for $i=0$, 1, 2 where $f^{(i)}(x)$ is the ith derivative of $f(x)$ at x. We formulate these constraints using matrix notation as

$$C^{\mathrm{T}}\beta = 0$$

where

$$C^{\mathrm{T}} \equiv \begin{pmatrix} 0 & \phi_1(0)-\phi_1(1) & \dots & \phi_K(0)-\phi_K(1) \\ 0 & \phi_1^{(1)}(0)-\phi_1^{(1)}(1) & \dots & \phi_K^{(1)}(0)-\phi_K^{(1)}(1) \\ 0 & \phi_1^{(2)}(0)-\phi_1^{(2)}(1) & \dots & \phi_K^{(2)}(0)-\phi_K^{(2)}(1) \end{pmatrix}.$$

Note that C is a $(K+1)\times 3$ matrix, and consider the singular value decomposition

$$C = USV^{\mathrm{T}} = (\,U_1 \quad U_2\,)\begin{pmatrix} S_1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} V_1^{\mathrm{T}} \\ V_2^{\mathrm{T}} \end{pmatrix}.$$

Here S_1 corresponds to the part of S with non-zero diagonal entries. Next, observe that $C^{\mathrm{T}}\beta = V_1 D_1 U_1^{\mathrm{T}}\beta$ equals zero if and only if $U_1^{\mathrm{T}}\beta = 0$. Using $UU^{\mathrm{T}} = I$, we can rewrite $f(x)$ as follows:

$$\begin{aligned} f(x) &= \Phi(x)^{\mathrm{T}}\beta = \Phi(x)^{\mathrm{T}}UU^{\mathrm{T}}\beta \\ &= (\,\Phi(x)^{\mathrm{T}}U_1 \quad \Phi(x)^{\mathrm{T}}U_2\,)\begin{pmatrix} U_1^{\mathrm{T}}\beta \\ U_2^{\mathrm{T}}\beta \end{pmatrix} = \Phi(x)^{\mathrm{T}}U_2U_2^{\mathrm{T}}\beta \\ &= \tilde{\Phi}(x)^{\mathrm{T}}\tilde{\beta}. \end{aligned}$$

Hence $f(x)$ is also the solution of an unconstrained regression using the transformed basis vector $\tilde{\Phi}(x) = \Phi(x)^{\mathrm{T}}U_2$. The computation of the coefficient vector $\tilde{\beta} = U_2^{\mathrm{T}}\beta$ is straightforward using this transformation.

References

[1] M. Allmen, C.R. Dyer, Cyclic motion detection using spatiotemporal surfaces and curves, in: International Conference on Pattern Recognition, 1990, pp. 365–370.

[2] K. Amaya, A. Bruderlin, T. Calvert, Emotion from motion, in: Proceeding of the Conference on Graphics Interface, 1996, pp. 222–229.

[3] O. Arikan, D. Forsyth, Interactive motion generation from examples, in: SIGGRAPH, 2002, pp. 438–490.

[4] P. Besse, J. Ramsey, Principal component analysis of sampled curves, Psychometrika 51 (1986) 285–311.

[5] M.J. Black, A.D. Jepson, A probabilistic framework for matching temporal trajectories: condensation-based recognition of gestures and expressions, in: H. Burkhardt, B. Neumann (Eds.), European Conferernce on Computer Vision, ECCV-98, LNCS-Series, vol. 140, Springer, Freiburg, Germany, 1998, pp. 909–924.

[6] A. Bobick, J. Davis, An appearance-based representation of action, in: International Conference on Pattern Recognition, vol. 1, 1996, p. 307.

[7] A.F. Bobick, J.W. Davis, The recognition of human movement using temporal templates, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (3) (2001) 257–267.

[8] A.F. Bobick, A.Y. Johnson, Gait recognition using static, activity-specific parameters, in: Proceedings of IEEE Computer Vision and Pattern Recognition Conference, vol. I, 2001, pp. 423–430.

[9] R. Bowden, Learning statistical models of human motion, in: IEEE Workshop on Human Modelling, Analysis and Synthesis, 2000.

[10] M. Brand, Shadow puppetry, in: International Conference on Computer Vision, vol. 2, 1999, pp. 1237–1244.

[11] M. Brand, A. Hertzmann, Style machines, in: SIGGRAPH, 2000, pp. 183–192.

[12] C. Bregler, Learning and recognizing human dynamics in video sequences, in: Proceedings of IEEE Computer Vision and Pattern Recognition, 1997, pp. 568–574.

[13] C. Bregler, J. Malik, Tracking people with twists and exponential maps, in: Proceedings of IEEE Computer Vision and Pattern Recognition, 1998, pp. 8–15.

[14] L.W. Campbell, A.F. Bobick, Recognition of human body motion using phase space constraints, in: Proceedings of IEEE Computer Vision and Pattern Recognition, Boston, MA, June 1995, pp. 624–630.

[15] T.-J. Cham, J.M. Rehg, A multiple hypothesis approach to figure tracking, in: Proceedings of IEEE Computer Vision and Pattern Recognition, vol. 1, 1999, pp. 239–245.

[16] R. Collins, R. Gross, J. Shi, Silhouette-based human identification from body shape and gait, in: Proceedings of IEEE Conference on Automatic Face and Gesture Recognition, 2002, pp. 351–356.

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

[18] R. Cutler, C. Benabdelkader, L.S. Davis, Motion based recognition of people in Eigen-Gait space, in: Proceedings IEEE Conference on Automatic Face and Gesture Recognition, 2002, pp. 267–272.

[19] J. Deutscher, I. Reid, Articulated body motion capture by stochastic search, International Journal of Computer Vision 61 (2) (2005) 185–205.

[20] J.J. Faraway, Regression analysis for a functional response, Technometrics 39 (1997) 254–261.

[21] M. Gleicher, Animation from observation: motion capture and motion editing, Computer Graphics 33 (4) (1999) 51–54.

[22] L. Goncalves, E. Di Bernardo, P. Perona, Reach out and touch space (motion learning), in: IEEE International Conference on Automatic Face and Gesture Recognition, 1998, pp. 234–238.

[23] N. Gordon, D. Salmond, A. Smith, A novel approach to nonlinear/non-Gaussian Bayesian state estimation, IEE Proceedings on Radar, Sonar and Navigation 140 (2) (1993) 107–113.

[24] T. Hastie, R. Tibshirani, G. Sherlock, M. Eisen, O. Alter, D. Botstein, P. Brown, Imputing missing data for gene expression arrays, Working Paper, Department of Statistics, Stanford University, 2000.

[25] D.C. Flogg, Model-based vision: a program to see a walking person, Image and Vision Computing 1 (1) (1983) 5–20.

[26] N.R. Howe, M.E. Leventon, W.T. Freeman, Bayesian reconstruction of 3D human motion from single-camera video, in: S.A. Solla, T.K. Leen, K.-R. Müller (Eds.), Advances in Neural Information Processing Systems, vol. 12, The MIT Press, Cambridge, MA, 2000, pp. 820–825.

[27] P.S. Huang, C.J. Harris, M.S. Nixon, Human gait recognition in canonical space using temporal templates, IEE Proceedings—Vision Image and Signal Processing 146 (2) (1999) 93–100.

[28] M. Isard, A. Blake, Contour tracking by stochastic propagation of conditional density, in: European Conference on Computer Vision, 1996, pp. 343–356.

[29] M. Isard, A. Blake, Condensation—conditional density propagation for visual tracking, International Journal of Computer Vision 29 (1) (1998) 5–28.

[30] S.X. Ju, M.J. Black, Y. Yacoob, Cardboard people: a parameterized model of articulated motion, in: IEEE International Conference on Automatic Face and Gesture Recognition, 1996, pp. 38–44.

[31] A. Kale, A. Sundaresan, A.N. Rajagopalan, N.P. Cuntoor, A.K. Roy-Chowdhury, V. Kruger, R. Chellappa, Identification of humans using gait, IEEE Transactions on Image Processing 13 (9) (2004) 1163–1173.

[32] L. Kovar, M. Gleicher, F. Pighin, Motion graphs, in: SIGGRAPH, 2002, pp. 491–500.

[33] F. De la Torre, M.J. Black, A framework for robust subspace learning, International Journal of Computer Vision 54 (1–3) (2003) 117–142.

[34] J. Lee, J. Chai, P. Reitsma, J. Hodgins, N. Pollard, Interactive control of avatars animated with human motion data, 2002, pp. 491–500.

[35] J.J. Little, J.E. Boyd, Recognizing people by their gait: the shape of motion, Videre: Journal of Computer Vision Research 1 (2) (1998) 1–32.

[36] L. Molina, A. Hilton, Realistic synthesis of novel human movements from a database of motion capture examples, in: IEEE Workshop on Human Motion, 2000, pp. 137–142.

[37] H. Murase, R. Sakai, Moving object recognition in eigenspace representation: gait analysis and lip reading, Pattern Recognition Letters 17 (1996) 155–162.
[38] M.P. Murray, Gait as a total pattern of movement, American Journal of Physical Medicine 46 (1) (1967) 290–332.
[39] M.S. Nixon, J.N. Carter, Advances in automatic gait recognition, in: Proceedings of IEEE Conference Automatic Face and Gesture Recognition, 2004, pp. 139–144.
[40] R.A. Olshen, E.N. Biden, M.P. Wyatt, D.H. Sutherland, Gait analysis and the bootstrap, The Annals of Statistics 17 (1989) 1419–1440.
[41] D. Ormoneit, H. Sidenbladh, M.J. Black, T. Hastie, Learning and tracking cyclic human motion, in: K.Leen Todd, G.Dietterich Thomas, Volker Tresp (Eds.), Advances in Neural Information Processing Systems, vol. 13, The MIT Press, Cambridge, MA, 2001, pp. 894–900.
[42] V. Pavolvić, J. Rehg, T.-J. Cham, K. Murphy, A dynamic Bayesian network approach to figure tracking using learned dynamic models, in: International Conference on Computer Vision, vol. 1, 1999, pp. 94–101.
[43] K. Pullen, C. Bregler, Animating by multi-level sampling, in: IEEE Computer Animation, 2000, pp. 36–42.
[44] K. Pullen, C. Bregler, Motion capture assisted animation: texturing and synthesis, in SIGGRAPH, 2002, pp. 501–508.
[45] J.O. Ramsey, B.W. Silverman, Functional Data Analysis, Springer, New York, 1997.
[46] P. Hall, J. Reimann, J. Rice, Nonparametric estimation of a periodic function, Technical report, University of California, Berkeley, 1999.
[47] J.A. Rice, B.W. Silverman, Estimating the mean and covariance structure nonparametrically when the data are curves, Journal of the Royal Statistical Society B 53 (1991) 233–243.
[48] K. Rohr, Towards model-based recognition of human movements in image sequences, CVGIP—Image Understanding 59 (1) (1994) 94–115.
[49] K. Rohr, Human movement analysis based on explicit motion models, in: M. Shah, R. Jain (Eds.), Motion-Based Recognition, Kluwer Academic Publisher, Dordrecht, 1997, pp. 171–198.
[50] S.M. Seitz, C.R. Dyer, Affine invariant detection of periodic motion, in: Proceedings of IEEE Computer Vision and Pattern Recognition, Seattle, WA, June 1994, pp. 970–975.
[51] H. Shum, K. Ikeuchi, R. Reddy, Principal component analysis with missing data and its application to polyhedral object modeling, IEEE Transactions on Pattern Analysis and Machine Intelligence 17 (9) (1995) 855–867.
[52] H. Sidenbladh, Probabilistic Tracking and Reconstruction of 3D Human Motion in Monocular Video Sequences, PhD Thesis, KTH, Stockholm, Sweden, 2001.
[53] H. Sidenbladh, M.J. Black, D.J. Fleet, Stochastic tracking of 3D human figures using 2D image motion, in: H. Burkhardt, B. Neumann (Eds.), European Conference on Computer Vision, LNCS-Series, vol. 1813, Springer, London, 2000, pp. 702–718.
[54] H. Sidenbladh, F. De la Torre, M.J. Black, A framework for modeling the appearance of 3d articulated figures, in: International Conference on Automatic Face and Gesture Recognition, Grenoble, France, March 2000, pp. 368–375.
[55] R. Tanawongsuwan, A. Bobick, Gait recognition from time-normalized joint-angle trajectories in the walking plane, in: Proceedings of IEEE Computer Vision and Pattern Recognition Conference, vol. II, 2001, pp. 726–731.
[56] N.F. Troje, Decomposing biological motion: a framework for analysis and synthesis of human gait patterns, Journal of Vision 2 (5) (2002) 371–387.
[57] O. Troyanskaya, M. Cantor, G. Sherlock, P. Brown, T. Hastie, R. Tibshirani, D. Botstein, R.B. Altman, Missing value estimation methods for DNA microarrays, Bioinformatics 17 (6) (2001) 520–525.
[58] M. Unuma, K. Anjyo,R. Takeuchi, Fourier principles for emotion-based human figure animation, in: SIGGRAPH, 1995, pp. 91–96.
[59] R. Urtasun, D.J. Fleet, P. Fua, Monocular 3D tracking of the golf swing, in: Proceedings of IEEE Computer Vision and Pattern Recognition, vol. II, 2005, pp. 932–938.
[60] R. Urtasun, P. Fua, 3D human body tracking using deterministic motion models, in: European Conference on Computer Vision, vol. III, 2004, pp. 92–106.
[61] S. Wachter, H.H. Nagel, Tracking persons in monocular image sequences, Computer Vision and Image Understanding 74 (3) (1999) 174–192.
[62] D.A. Winter, The Biomechanics and Motor Control of Human Gait: Normal, Elderly and Pathalogical, second ed., Waterloo Biomechanics, Waterloo, Ont., 1991.
[63] C.R. Wren, A.P. Pentland, Dynaman: recursive modeling of human motion, Technical Report TR-451, MIT, 1999.
[64] H.-Y. Shum, Y. Li, T. Wang, Motion texture: a two-level statistical model for character motion synthesis, in: Proceedings of ACM SIGGRAPH, 2002, pp. 465–472.
[65] Y. Yacoob, M.J. Black, Parameterized modeling and recognition of activities in temporal surfaces, Computer Vision and Image Understanding 73 (2) (1999) 232–247.
[66] Y. Yacoob, L. Davis, Learned models for estimation of rigid and articulated human motion from stationary or moving camera, International Journal of Computer Vision 36 (1) (2000) 5–30.