Reconstructing and analyzing periodic human motion from stationary monocular views

Evan Ribnick, Ravishankar Sivalingam, Nikolaos Papanikolopoulos, and Kostas Daniilidis

Abstract
We have shown previously that it is possible to accurately reconstruct periodic motions in 3D from a single camera view, using periodicity as a physical constraint from which to perform geometric inference. In this paper we explore the suitability of the reconstruction techniques for real human motion. We examine the degree of periodicity of human gait empirically, and develop algorithmic tools to address some of the challenges arising from this type of motion, including reconstructing motions that deviate from pure periodicity, properly handling the trajectories of multiple points on an articulated body, and proposing a distance function for measuring the difference between two reconstructions. Importantly, we illustrate the usefulness of these techniques by applying them to the tasks of view-invariant activity classification, clinical gait analysis and person identification.

1. Introduction
Periodicity has been recognized as an important cue in many computer vision tasks, especially those involving the analysis of human motion. For example, Briassouli and Ahuja [1] use Fourier analysis to segment and extract multiple simultaneous periodic motions in video sequences. However, much of the existing research presents techniques which operate only in image coordinates, and are therefore heavily dependent on the viewing angle. This can limit their applicability for tasks like motion analysis, recognition, and classification, since the appearance of a motion in image coordinates will vary greatly with the viewing angle – see, for example, Fig. 1. Specifically, these algorithms may not generalize well to viewing angles not seen in the training data.

We have shown in our previous work [2–4] that a feasible alternative exists, in which a motion that is periodic can be reconstructed in 3D based on its image-coordinate trajectory in a single camera view. In our formulation periodicity is used as a physical constraint, and algorithms for effectively performing geometric inference are introduced, as well as theoretical results indicating that such reconstructions are possible to obtain. The idea is that, once a periodic trajectory has been reconstructed, any subsequent analysis can be performed in 3D world coordinates, independent of the original viewing angle. It was shown that these techniques can be accurate, robust, and view-invariant in practice.

Due to these challenges, additional algorithmic tools are needed in order to effectively analyze these types of movements, and a more complete understanding of the periodicity exhibited by natural human motion is required.

In this paper we present a framework for reconstructing and analyzing real periodic human motion, including the surrounding issues that arise with the complexities of this domain, in which we attempt to answer all of the questions listed above. Question

1. How periodic is real human motion?
2. Is it possible to reconstruct repetitive motions that deviate significantly from pure periodicity?
3. What is an appropriate way to handle the trajectories of multiple points on an articulated body?
4. What kinds of distance functions should be used to robustly compare reconstructed trajectories?
1 is examined empirically by carefully studying motion capture data of real human gait (Section 7). As a result, we find that human motion can adhere rather closely to pure periodicity over short durations of time, indicating the feasibility of applying our reconstruction techniques in this domain.

Nonetheless, question 2 refers to repetitive motion that is not well explained by pure periodicity, such as that of a person who changes speed while walking. This is addressed here by developing an algorithm for reconstructing periodic motions whose temporal period varies over time (Sections 5 and 8). Likewise, question 3 is also important, since in some cases it may be necessary to use the motion of multiple points on the human body in order to analyze the properties of its overall movement. We answer this by presenting a technique for linking the reconstructions of multiple points on an articulated body using a set of rigidity constraints (Section 4). Finally, question 4 reminds us that before these reconstructions can be applied for any particular task, it will be necessary to introduce a distance function for computing differences between trajectories – this is addressed here in Section 6.

The applicability of these reconstruction techniques to real human motion is also explored through pilot studies demonstrating their usefulness in three different areas. In the first (Section 9), we use the proposed algorithms for the task of activity classification, which has the additional advantage of being view-invariant. Next, Section 10 illustrates the applicability to clinical gait analysis, in which we successfully distinguish between the gaits of people who have been partially debilitated by a stroke, and those of healthy individuals. Lastly, Section 11 tackles the challenging task of identifying different people based on their characteristic gait. In all cases we illustrate the usefulness of the proposed reconstruction techniques for real human motion, and show that a significant amount of information can be gleaned from the trajectories of a very small number of points on the body.

Note that the focus of this work is on the specifics of tracking points of interest – instead, we choose to focus on the issues involved with reconstructing and analyzing periodic human motions in 3D. As such, it is assumed here that image-coordinate tracks are available. In some cases they have been obtained by tracking a brightly colored marker in the image. It is important to note that, in many application areas, including those in which we are interested (see Section 10 on clinical gait analysis), the use of marker-based tracking is the norm and is completely acceptable, since analysis is performed in a controlled environment. This is in contrast to other applications such as surveillance, in which the subject may be less than cooperative. Furthermore, a similar assumption was made in the closely related work [5], in which a commercial motion-capture system was used to obtain accurate tracks in 3D (in fact, we operate under fewer assumptions here).

2. Related work

A large majority of the existing work related to periodic motion has focussed on detection and analysis in image coordinates, and does not attempt to extract the 3D information embedded in these motions. For example, several techniques use Fourier analysis of pixel-coordinate movements to detect, segment, or classify motions – see [6–10,11–13]. Other work, such as [14–17], has used other types of image-based analysis for detection, representation, and classification of repetitive motions, including human gestures and facial expressions.

There is, however, a small kernel of research which has recognized the connection between periodic motion and 3D inference. As mentioned before, we make use here of our prior work [2–4], which presented algorithms for inferring the 3D trajectory of periodic motions based on their image-coordinate trajectories and geometric constraints. Similarly, Belongie and Wills [18] propose a method for estimating the structure of an articulated body which is undergoing repetitive motion by considering snapshots separated by exactly one period in time, and also use geometric constraints. Zhang and Troje [5] propose another technique, in which training data is used to learn Fourier-based representations of periodic human motions, which can be used to infer their structure in 3D. This is different from our approach, since its applicability may be limited by the trajectories seen in training data, and an orthographic projection is assumed, indicating reconstruction only for motions without a translational component.

More broadly, this work is also related to the general area of vision-based human motion analysis – in fact, we consider the task of activity classification in Section 9, which is one of the common objectives in this domain. There is an extremely large amount of activity in this field, and we only cite a few representative papers here – for more thorough coverage, the reader is referred to the survey papers [19–22]. The research in this area can generally be divided into two major categories – those that perform analysis from a single camera view (monocular systems), and those that require multiple views (e.g., binocular and trinocular stereo). Most relevant to this paper is the first category, monocular systems. Typically they either operate directly in image coordinates [23], or learn mappings from image-coordinate appearances to 3D state vectors [24]. In any case, these techniques are affected by the viewing angle of the camera with respect to the actor, and can only hope to overcome this constraint by learning classifiers based on motions viewed from many different angles.

Finally, Section 10 examines gait analysis in a clinical setting, so we also consider related work in that domain. The very important and classical work of [25] gives a thorough characterization of the gait of adult men, and clearly illustrates the connection between a person’s gait and the presence/absence of physical debilitations. Otherwise, the amount of literature dealing with technology-assisted gait analysis for medical diagnosis is rather sparse. In [26], the authors also show that gait is a useful cue for diagnosis of pathological disorders, specifically cerebral palsy and poliomyelitis. The signals of interest are the motion of the lower extremities. Fourier analysis is performed on the resulting signals, and Kohonen maps (i.e., self-organizing maps) are used to cluster the data. It is shown that the data from normal and pathological subjects are well separated. Similarly, [27] develop a system for diagnosing Parkinson’s disease. The system analyzes static images of the walking person wearing a conveniently colored track suit, and an artificial neural network is trained in order to distinguish between diseased and normal subjects based on several features extracted from the

Fig. 1. Two simultaneous views of a person walking. Even though the exact same motion is viewed in both cases, the appearance of the trajectories in image coordinates varies significantly with viewing angle.
images. Other works, such as [28–30] among others, use commercial motion capture systems to analyze gait.

3. Periodic motion reconstruction

In this section we briefly review the basic formulation and technique for reconstructing periodic motions in 3D. For more detailed discussion, the reader is referred to [3,4].

3.1. Definitions and basic equations

Periodic motion is defined in this work as any movement that is periodic in velocity (in 3D world coordinates):

\[ \mathbf{v}(t + nT) = \mathbf{v}(t), \]

for any integer \( n \), where \( \mathbf{v} = (X, Y, Z) \) and \( T \) is the period. In terms of the 3D position of the point, we have:

\[ \mathbf{p}(t + nT) = \mathbf{p}(t) + n\Delta \mathbf{p}, \]

where \( \Delta \mathbf{p} = (\Delta X, \Delta Y, \Delta Z) \) is the displacement per period of the point, which is constant over any period of length \( T \). For example, if the point being tracked is on the foot of a walking person, then the stride length is equal to \( |\Delta \mathbf{p}| \).

Since samples are taken at discrete times determined by the video frame rate, we represent times using discrete indices of the form \( t_k \). This represents the time of the \( k \)th sample in the \( k \)th period. We can then arrive at the following expression for the position at time \( t_k \):

\[ \mathbf{p}_k = \mathbf{p}_k^0 + i\Delta \mathbf{p}, \]

which is written in expanded form as:

\[ \begin{bmatrix} X_k^i \\ Y_k^i \\ Z_k^i \end{bmatrix} = \begin{bmatrix} X_k^0 \\ Y_k^0 \\ Z_k^0 \end{bmatrix} + i \begin{bmatrix} \Delta X_k \\ \Delta Y_k \\ \Delta Z_k \end{bmatrix}, \]

where the sample at time \( t_k \) is expressed relative to \( t_0^i \), the \( 0 \)th sample in the zeroth period.

When a periodic motion is projected into the image using the pinhole camera model, we arrive at the pixel-coordinate trajectory described by the following equation:

\[ \begin{bmatrix} u_k^i \\ v_k^i \end{bmatrix} = \frac{1}{Z_k^0 + i\Delta Z_k} \begin{bmatrix} f_x & 0 \\ 0 & f_y \end{bmatrix} \begin{bmatrix} X_k^0 + i\Delta X_k \\ Y_k^0 + i\Delta Y_k \end{bmatrix} + \mathbf{c}_x + i\mathbf{c}_y. \]

Note that we have placed the origin of the world coordinate system at the camera center, with the Z-axis parallel to the camera’s optical axis. The quantities \( f_x, f_y, c_x, \) and \( c_y \) are intrinsic parameters of the camera representing the focal length and image plane center in pixels. Thus, we see that it is possible to express the projection of any sample projected into image coordinates, \( (u_k^i, v_k^i) \), as a function of the corresponding 3D sample at time \( t_k^i \) and the inter-period displacement \( \Delta \mathbf{p} \).

3.2. Minimizing the 3D geometric error

Rearranging the terms in (5), we can obtain expressions for \( X_k^0 \) and \( Y_k^0 \) in terms of estimates of \( Z_k^0 \) and \( (\Delta X_k, \Delta Y_k, \Delta Z_k) \):

\[ i\hat{X}_k^0 = \frac{u_k^i - c_x}{f_x} \left( \hat{Z}_k^i + i\Delta Z_k \right) - i\hat{\Delta}X_k, \]

\[ i\hat{Y}_k^0 = \frac{v_k^i - c_y}{f_y} \left( \hat{Z}_k^i + i\Delta Z_k \right) - i\hat{\Delta}Y_k, \]

where \( \hat{\cdot} \) denotes that a quantity is an estimate, and \( \hat{X}_k^0 \) and \( \hat{Y}_k^0 \) are approximations of \( X_k^0 \) and \( Y_k^0 \) based on the estimates and the image-coordinate samples of period \( i \). Such equations can be formed for each sample \( k \) and each period \( i = 0, 1, \ldots, M - 1 \).

Ideally \( \hat{X}_k^0 = \hat{Y}_k^0 = \hat{X}_k \) and \( \hat{Y}_k^0 = \hat{Y}_k \) for any sample \( k \) and any pair of periods \( i_1 \) and \( i_2 \). Therefore, making use of (6) and (7), we can obtain a pair of equations as follows:

\[ \begin{align*}
  i\hat{X}_k^{i_1} - i\hat{X}_k^{i_2} &= \frac{u_k^{i_1} - u_k^{i_2}}{f_x} \hat{Z}_k^{i_1} + (i_2 - i_1)\hat{\Delta}X_k \\
  &+ \frac{1}{f_x} (i_1 u_k^{i_1} - i_2 u_k^{i_2}) + c_x (i_2 - i_1) \hat{\Delta}Z_k = 0 \\
  i\hat{Y}_k^{i_1} - i\hat{Y}_k^{i_2} &= \frac{v_k^{i_1} - v_k^{i_2}}{f_y} \hat{Z}_k^{i_1} + (i_2 - i_1)\hat{\Delta}Y_k \\
  &+ \frac{1}{f_y} (i_1 v_k^{i_1} - i_2 v_k^{i_2}) + c_y (i_2 - i_1) \hat{\Delta}Z_k = 0. \end{align*} \]

Two equations of the form (8) and (9) can be obtained for every sample \( k \), for every pair of periods \( i_1 \) and \( i_2 \). This results in a total of \( 2N M \), where \( M \) is the number of periods, and \( N \) is the number of samples from each period. If we stack all these equations together in matrix form, the result is an overconstrained homogeneous linear system, which can be solved by performing the following optimization:

\[ \text{minimize} \quad ||AX||_2, \]

subject to \( ||X||_2 = 1 \).

where \( A \) is the coefficient matrix, and:

\[ X = \begin{bmatrix} \hat{Z}_0^0 & \hat{Z}_1^0 & \hat{Z}_2^0 & \ldots & \hat{Z}_{N-1}^0 & \hat{\Delta}X_0 & \hat{\Delta}X_1 & \hat{\Delta}X_2 & \ldots & \hat{\Delta}X_{N-1} & \hat{\Delta}Y_0 & \hat{\Delta}Y_1 & \hat{\Delta}Y_2 & \ldots & \hat{\Delta}Y_{N-1} \end{bmatrix}^T. \]

The optimization (10) can be performed efficiently using one Singular Value Decomposition (SVD), where \( A = UV \), and the minimizer \( X \) is the last column of \( V \).

3.3. Constant period estimation

The period of motion is also be found in order to perform reconstruction, and can be estimated using Fourier analysis of the image-coordinate velocity signals, \( \mathbf{u}(t) \) and \( \mathbf{v}(t) \). Specifically, different linear combinations of the signals of the form:

\[ \nu_p(t) = \begin{bmatrix} \cos(\phi) \\ \sin(\phi) \end{bmatrix} \frac{1}{T} \begin{bmatrix} \mathbf{u}(t) \\ \mathbf{v}(t) \end{bmatrix}. \]

parameterized by the angle \( \phi \), are computed. The Power Spectral Density (PSD) of each signal is calculated, and all the PSDs are superimposed with weights proportional to their sparsity (as given by the \( \ell = 0.5 \) norm). Finally, the period of motion is taken as the inverse of the largest spectral component in this weighted sum of PSDs.

4. Reconstructing multiple points on an articulated body

In this work we are interested in the natural periodic motion of the human body when it performs common activities such as walking and running. To analyze these types of motion, it will be necessary to compare reconstructed trajectories by computing distances between them. In some cases, we will track multiple points on the body as a motion is performed, which can yield more information than just a single point alone. These points will typically be on the skeletal joints, so that pairs of points are rigidly connected by the bones between them (for example, such is the case with points on the ankle and knee). As such, large numbers of points on an articulated body can be thought of as connected by pairs of rigidity constraints. However, when comparing reconstructions of multiple body points between different people, it is important
to ensure that the rigid connection between pairs of points are maintained in order for the comparison to be valid.

As an example, consider the situation depicted in Fig. 2, in which we see two examples of a stick (which could represent, for example, a lower leg) moving along a periodic trajectory. If one were to consider the trajectories of only the upper endpoints of the line segments (shown in green), it would be seen that they are identical in these two cases, up to a translation. The same will hold for the trajectories of the lower endpoints (shown in blue). However, if the motion of the line segment is viewed as a whole, it is clear that the two cases are actually quite different due to the angle of the stick. As such, we see that the rigidity constraints linking two point trajectories must be enforced in order to make accurate comparisons, which take the full scenario into account.

One way to enforce these rigidity constraints would be to reconstruct the trajectories of all connected points simultaneously using the standard single-point reconstruction technique for periodic motion, along with some additional constraints enforcing the pairwise rigidity between points. However, as we will see momentarily, the equation for a rigidity constraint is very nonlinear in the variables which we have used to parameterize a periodic motion, which would result in a non-convex cost function that is difficult to optimize reliably. Instead, we adopt a strategy in which the rigidity constraints are enforced as an additional step, after reconstructing each of the point trajectories independently.

Let \( \mathbf{p}_i \) and \( \mathbf{q}_i \) represent the 3D trajectories of two rigidly connected points, where as before \( k = 0, 1, \ldots, N - 1 \) are the discrete time indices (for simplicity we omit the period index \( n \) momentarily). Assume that these two point trajectories have been reconstructed using the techniques described in Section 3. Since each reconstruction is accurate up to a scale factor, it will be sufficient to scale \( \mathbf{q}_i \) such that each of its points is a uniform distance from the corresponding point in \( \mathbf{p}_i \). If \( d \) is the distance between the rigidly connected points, then for each time \( t_i \) we have:

\[
\| \mathbf{p}_i - \lambda \mathbf{q}_i \|_2^2 = d^2,
\]

where \( \lambda \) is a scale factor. However, the quadratic Eq. (13) might not have a real solution depending on the value of \( d \), so we instead write the constraint as follows:

\[
\| \mathbf{p}_i - \lambda \mathbf{q}_i \|_2^2 = \| \mathbf{p}_{i+1} - \lambda \mathbf{q}_{i+1} \|_2^2
\]

or equivalently:

\[
\lambda^2 (q_{i,3} - q_{i,0})^2 + \lambda (2q_{i,1} - 2q_{i,0}) + (p_{i,3} - p_{i,0}) = 0
\]

for every pair of time indices \( t_i \) and \( t_{i+1} \). If all constraints of the form Eq. (15) are stacked together in vector form, we obtain the following overall equation:

\[
\lambda^2 \mathbf{c}_1 + \lambda \mathbf{c}_2 + \mathbf{c}_3 = \mathbf{0},
\]

where the vectors \( \mathbf{c}_1, \mathbf{c}_2, \) and \( \mathbf{c}_3 \) are the result of stacking together the coefficients in (15) in vector form.

In the ideal, noise-free case, there is a scale factor \( \lambda \) such that (16) is satisfied. However, in practice the constraints will be slightly violated due to noise and other sources of error, so we attempt to find the best least-squares solution:

\[
\lambda = \arg \min \| \lambda^2 \mathbf{c}_1 + \lambda \mathbf{c}_2 + \mathbf{c}_3 \|_2^2.
\]

Fortunately, (17) expands to a fourth-order polynomial which can be minimized in closed form by finding the roots of the derivative \( f'\lambda \) and choosing the one that minimizes \( f(\lambda) \).

Finally, note that this formulation can easily be extended to include a larger number of tracked points linked by a set of rigidity relationships, such as when several joints on an articulated body are tracked simultaneously.

5. Reconstruction with non-constant period

The periodic motion reconstruction technique described in Section 3 assumes that the period of the motion in world coordinates is constant, since equations are formed from pairs of samples at corresponding phases in different periods. If this assumption is violated, the resulting reconstruction will be inaccurate.

Since some types of cyclic motion may not adhere closely to strict periodicity, it is necessary to develop a technique for reconstructing this more general class of motions. We focus our attention on the case where only the instantaneous period (or equivalently, the instantaneous fundamental frequency) fluctuates over time. In other words, the shape of the path followed by the point in 3D remains the same, but the speed at which it is traversed varies. This may be a reasonable approximation for many types of cyclic motion.

We will use the concept of period trace to describe the temporal fluctuations in periodicity. The period trace was introduced by [14], and is an elegant way of modeling the instantaneous period of a cyclic motion. It can be defined as follows.\(^1\) First, we define \( \phi(t) \) as the unwrapped phase of \( \phi(t) \) of the cyclic motion, which is a function of time. If the motion were perfectly periodic with period \( T \), then

\(^1\) Although the period trace was introduced by Seitz and Dyer, the definition presented here is different in that it is based on the unwrapped phase function \( \phi(t) \). This is not to be confused with the function \( \phi(t) \) in [14], which has a completely different meaning.
it would be true that $\phi(t) = \frac{2\pi}{T} t$, since one period encompasses a phase change of $2\pi$. In the more general case of cyclic motion, however, $\phi(t)$ can be any monotonically increasing function. Then the period trace at each time $t$ is:

$$
\tau_1(t) = \{T : \phi(t + T) = \phi(t) + 2\pi\},
$$

(18)

which indicates the time difference between the corresponding phase in the next period and the current time. For a periodic motion, $\tau_1(t)$ is constant and equal to the period.

As indicated in [14], a cyclic motion can be warped into a periodic motion by reparameterizing the time index such that the period trace becomes flat. In other words, if the function $f(t)$ is cyclic, then there is a time warping function $\mathcal{s}(t)$ such that $f(\mathcal{s}(t))$ is periodic. There exist many such reparameterizations $\mathcal{s}(t)$, but for our purposes it is computed as follows. First, the last period of motion is frozen and taken as the reference period, so that the aim is then to warp all other periods to match it. The constant period is chosen as the length of the reference period:

$$
T = \{\tau_1(t_0) \mid \tau_1(t_0) + t_0 = t_{\text{end}}\},
$$

(19)

where the signal $f(t)$ is of finite duration, and defined over the range $t \in [0, t_{\text{end}}]$. Finally, the reparameterization is defined as:

$$
\mathcal{s}(t) = \begin{cases} 
\frac{\mathcal{s}(t + \tau_1(t)) - T}{\tau_1(t)} & 0 \leq t < t_0 \\
\tau_1(t) & t_0 \leq t \leq t_{\text{end}}.
\end{cases}
$$

(20)

We can see from (20) that $\mathcal{s}(t)$ can only be computed if $\mathcal{s}(t + \tau_1(t))$ is already known. For this reason, we have chosen the last period of motion as the reference, so that in practice it is possible to proceed backwards through time to compute $\mathcal{s}(t)$ at each discrete time step. Note that interpolation becomes necessary in practice, since the time indices are discrete in any computerized implementation, and $t + \tau_1(t)$ may be non-integer in general. The resulting signal $f(\mathcal{s}(t))$ is defined over the range $t \in [\mathcal{s}(\tau_1(0)) - T, t_{\text{end}}]$.

5.1. Period trace estimation

The period trace must have the following properties in our particular case:

1. $p(t + \tau_1(t)) = p(t) + \Delta_p$, where $p(t)$ is the position of the point in 3D, and $\Delta_p$ is the inter-period displacement
2. $\tau_1(t)$ is continuous
3. $\tau_1(t) > 0$
4. $\tau_1(t) > -1$
5. $\tau_1(t)$ is defined in the range $t \in [0, t_0]$, where $\tau_1(t_0) + t_0 = t_{\text{end}}$.

If it were possible to observe the cyclic motion $p(t)$ directly or to make similarity measurements, the period trace could be estimated by fixing the last period as the reference period, and choosing $\tau_1(t)$ at each discrete time step such that the constraints above are satisfied according to the algorithm proposed in [14]. Unfortunately, however, $p(t)$ is not directly observable, and only its projection into image coordinates is known a priori.

Instead, we must estimate the period trace iteratively, where each iteration makes perturbations of the period trace, reconstructs according to each perturbation, and chooses the one that results in the smallest cost $F(\tau_1)$. The cost is defined as follows:

$$
F(\tau_1) = \|A_p X\|_2^2 + 2\sum_{t} \|\mathcal{A}_t(\tau_1)\|,
$$

(21)

where the second term is a regularizer which seeks to enforce smoothness in the period trace, and the first term is the reconstruction error. In order to reconstruct a cyclic motion according to a period trace $\tau_1(t)$, the image coordinate trajectory is first dewarped as described above, and then resampled at uniform discrete time steps (we use non-parametric Gaussian Process regression – see [32,33]).

In terms of the necessary properties of the period trace, the first term in the cost (21) can be thought of as enforcing property 1, while the second term enforces property 2. Properties 3 and 4 are enforced locally during the iterations.

The algorithm makes several passes through the period trace $\tau_1(t)$, and at each iteration of a pass, the period trace is perturbed at one discrete time sample $t_p$. The period trace is initially perturbed by adding $-1$, $0$, and $1$ to $\tau_1(t_p)$, and choosing the perturbation resulting in the smallest cost (21). Towards the end, when the algorithm is closer to convergence, the step size is reduced progressively from 1 to smaller fractions to obtain smoother period trace estimates. Additionally, a multiscale coarse-to-fine optimization strategy is employed to avoid local minima. To summarize, at each iteration in each pass, and for each scale in the coarse-to-fine sampling, the following steps are performed:

1. Perturb the period trace at $\tau_1(t_p)$, ensuring that the perturbations do not violate properties 3 and 4 above.
2. For each perturbed period trace $\tau_1(t)$, warp the image coordinate trajectory of the cyclic motion according to the reparameterization (20).
3. Resample the dewarped image coordinate trajectory using Gaussian Process regression.
4. Reconstruct the dewarped trajectory by minimizing $\|A_p X\|_2$ subject to $\|X\|_2 = 1$ as described previously.
5. Compute the cost of this period trace perturbation according to (21).
6. Choose the period trace that results in the lowest cost $F(\tau_1)$.

Finally, after convergence, the period trace estimate is further denoised using quadratic smoothing [34]. If we represent the noisy signal as a vector $\mathbf{X}$ consisting of the concatenation of the estimated period trace values at the discrete time samples, then quadratic smoothing attempts to minimize the following cost function to obtain the denoised estimate $\hat{\mathbf{X}}$:

$$
\|\hat{\mathbf{X}} - \mathbf{X}\|_2^2 + \delta\|D\hat{\mathbf{X}}\|_2^2,
$$

(22)

where $D$ is a coefficient matrix such that $D\mathbf{X}$ approximates, for example, the first derivative of the signal using forward differences. The parameter $\delta$ is a weighting coefficient. From (22) we can see that the goal of quadratic smoothing is to construct a signal that is as close (in the least-squares sense) to the original signal, but which is also smooth – in essence, this is a bi-criterion optimization problem. The cost function (22) has a closed-form solution:

$$
\hat{X}_1 = (I + \delta D^T D)^{-1} X_1.
$$

(23)

6. Trajectory distance functions

As discussed previously, 3D reconstructions of periodic motions are only accurate up to a scale factor in general. Additionally, if a single trajectory is reconstructed from two different views, it is difficult to compare the reconstructions directly, since they are each represented on a camera-centered coordinate system with an unknown geometric transformation between them. For these reasons, it becomes necessary to introduce a robust distance metric for comparing 3D reconstructions.

One of the difficulties in defining a distance metric is that the two reconstructions being compared might have different lengths (i.e., composed of different numbers of samples). One approach is to warp the trajectories to match as closely as possible by effectively reparameterizing the time index – a prime example of this is Dynamic Time Warping [35]. However, in the types of applications that we are interested in (i.e., human motion analysis), the period of motion may be an important cue. Furthermore, we prefer

...
Given this intended application area, it is then natural to ask the following question:

**How periodic is real human motion?**

In other words, one might be inclined to wonder about just how applicable these reconstruction techniques are when faced with natural human motion, which may not strictly adhere to any parametric motion model. Here we explore this question in detail in the form of quantitative analysis, and provide some interesting results.

The data analyzed in these experiments consists of gait sequences collected using a commercial motion-capture system. In each sequence, an infra-red marker was placed on the subject’s foot (in this case near the toe), and the motion-capture system returned accurate 3D trajectories for this point of interest. The data set contains 16 gait sequences from a single subject, with each gait sequence comprised of approximately 3–4 full strides. An example of one image from this data set is shown in Fig. 3.

It is important to note that this data illustrates the case of a cooperative subject who has been instructed to simply walk as evenly as possible, following a straight path between two points. It is clear that data collected in a more unstructured setting, in which the subject is uncooperative or even adversarial (for example, video from a surveillance camera) may contain more significant fluctuations. However, we have observed qualitatively that even in such situations, relatively constant periodicity is often exhibited over short lengths of time. In any case, even the most cooperative subject will certainly deviate somewhat from pure periodicity, since it would be impossible to follow a strict motion model to the degree measurable by such precise equipment.

### 7.1. Temporal periodicity

First we examine the temporal periodicity of these gait sequences – specifically, the constancy of the temporal duration of each stride of the gait. This is done by manually demarcating the beginning of each stride in every sequence. We find that, over all 16 gait sequences, the mean period is 2.46 s, with a standard deviation of 0.11 s (4.6% of the mean). Additionally, the average absolute change in period length between two successive strides is 0.06 s. In other words, the deviation in period between two successive strides is on average 0.06 s.

### 7.2. Spatial periodicity

Next we examine the spatial periodicity of these gait sequences. This includes the constancy of both the stride length and direction. As before, we make use of manual demarcations of individual strides – in other words, the location of the same phase in each
period has been indicated manually for these sequences. Over all 16 sequences, we find that the mean stride length is 106.1 cm, with a standard deviation of 7.53 cm (7.1% of the mean). The average absolute change in stride length between two successive strides was found to be 2.69 cm. In addition, the average absolute change in stride direction between two successive strides is 2.17°.

### 7.3. Inter-period prediction error

According to our definition (Section 3), periodic motion can be described as $p(t + T) = p(t) + \Delta p$. In other words, the displacement between any two samples separated by exactly one period in time is a constant factor, referred to as the inter-period displacement. In this experiment we examine the suitability of this motion model, which we have used in the derivation of our reconstruction techniques, to natural human motion. For each sequence, $\Delta p$ is first estimated by computing the mean displacement between each pair of samples separated by one period in time. Then, the position of each sample $p(t + T)$ is predicted as $p(t + T) = p(t) + \Delta p$, and the error between the prediction $\hat{p}(t + T)$ and the actual $p(t + T)$ is measured. This is sort-of a composite measure, which captures the overall periodicity of each gait sequence and measures the applicability of our motion model. Over all 16 gait sequences, we find that the average prediction error computed in this manner is 9.63 cm.

### 8. Reconstructing cyclic human motion

We have shown (Section 7) that, on average, human motion does adhere relatively well to the model of strict periodicity. In this experiment we analyze the case when this assumption is violated, so that the motion is more aptly referred to as cyclic rather than periodic, and compare the reconstruction accuracies of the constant and non-constant period formulations.

The data used in this experiment consisted of one of the motion-capture sequences from the set described in Section 7. From those 16 gait sequences, we chose the one whose temporal period varied the most according to the manual demarcations – in other words, the sequence that most severely violated pure periodicity in the temporal dimension. The 3D motion-capture track was then synthetically projected into image coordinates using the full perspective projection described earlier in order to simulate an image-coordinate track.

Reconstruction was performed using both the iterative non-constant period formulation (Section 5), as well as the standard geometric error formulation (Section 3), which assumes a constant period. Since the non-constant period formulation is iterative, the runtime was orders of magnitude longer than the standard formulation.

The period trace estimated by the iterative procedure is shown in Fig. 4 – note that this shows significant deviations from a constant period. Quantitatively, the results can be summarized in terms of reconstruction errors by aligning them with the motion capture data using the Orthogonal Procrustes Distance (OPD). We find that the non-constant period technique achieves a reconstruction error formulation (Section 3), which assumes a constant period. Since these techniques use only geometric constraints, they do not require any training data. Furthermore, we have already demonstrated [3] that it is possible to achieve view-invariance using these techniques, even if the motion is viewed from only a single camera. Here we use a simple application of the proposed algorithms to demonstrate the possibility of activity classification in a view-invariant fashion.

Specifically, in this experiment we show that the reconstructed trajectory of a single point on the body (near the ankle in this case) can be used to distinguish between several activities. It is important to note that the aim of this work is not to propose a new method for activity recognition. Rather, we choose the task of activity classification only to demonstrate that useful information is embedded in these reconstructions.

### 9. Application: activity classification

In this section we demonstrate the applicability of the proposed techniques for reconstructing periodic motions to the task of activity classification. Of particular interest is the possibility of classifying activities in a view-invariant manner.

The appeal of the proposed methodologies for use in activity classification stems from the fact that many common activities, such as walking and running, are inherently repetitive in nature. Since these techniques use only geometric constraints, they do not require any training data. Furthermore, we have already demonstrated [3] that it is possible to achieve view-invariance using these techniques, even if the motion is viewed from only a single camera. Here we use a simple application of the proposed algorithms to demonstrate the possibility of activity classification in a view-invariant fashion.

Specifically, in this experiment we show that the reconstructed trajectory of a single point on the body (near the ankle in this case) can be used to distinguish between several activities. It is important to note that the aim of this work is not to propose a new method for activity recognition. Rather, we choose the task of activity classification only to demonstrate that useful information is embedded in these reconstructions.

#### 9.1. Data collection

We analyzed videos containing two different people performing several instances of the following four activities: walking, jogging, marching, and sideways shuffling. Motion sequences were filmed from two very different viewing angles. Examples of these four activities are shown in Fig. 5 from the two different views. In total, the dataset consisted of 32 sequences, including four instances of each activity from each of the two views. The subjects were asked only to perform the given action as evenly as possible, and to travel between two designated points. In each sequence, the point of interest was near the subject’s ankle, and was tracked using a brightly colored marker attached to the subject’s clothing.

#### 9.2. Trajectory reconstruction

Once the points of interest had been tracked in image coordinates as described above, the period of motion was estimated for each sequence using the technique described in Section 3.3. Each sequence was then reconstructed using the 3D geometric error cost function (Section 3.2). We used the constant period formulation for this experiment, since the sequences used here did adhere closely to the model of strict periodicity.
9.3. Distance computations

Before we can compute distances between the reconstructed activity sequences, it is first necessary to scale them in a consistent way, so that results obtained from any subsequent distance computations will be scale-invariant. To this end, all sequences are initially interpolated and resampled to have the same period. One period of one reconstruction is then designated as the reference period. Next, for each sequence, every one-period segment of ankle trajectory is scaled, rotated, and translated so that it is best-aligned with the reference period using the Orthogonal Procrustes Distance (OPD) (Section 6). For the one-period segment that most closely matches the reference period, the scale-factor computed for that alignment is then used as the overall scaling of that sequence. The computed scale factor for each sequence is finally applied to the original reconstruction (i.e., not the resampled reconstruction).

After scaling, the distance between each pair of reconstructions is calculated using the OPD, which computes the rotation and translation that minimizes the Mean Square Error (MSE) between them. Note that the distance between a pair of trajectories is taken as the MSE at the temporal overlap at which the minimum distance is achieved – in other words, we choose the temporal alignment of the sequences which minimizes the MSE, as described in Section 6.

9.4. Sequence classification

Each sequence is classified as belonging to one of the four activities using an opposite-view nearest-neighbor classification scheme, in which a reconstructed trajectory from one view is assigned an activity label based on its k-nearest-neighbors from the reconstructed trajectories from the other view. During classification, the sequence being classified is treated as the test data, while all reconstructed sequences from the opposite view comprise the training data. This opposite-view classification was chosen to further highlight the view-invariant nature of these reconstructions.

Results from the opposite-view activity classification are shown in Table 1 in the form of a confusion matrix. Overall classification accuracy is 96.88% – only one of the 32 trajectories was misclassified. Recall that this activity classification performance was achieved using only one tracked point on the body.

9.5. Manifold embedding

We further explore the information content of these gait reconstructions by examining their embedding in a low-dimensional manifold space. Each gait sequence was taken as a point on the manifold in some high-dimensional space, and embeddings were computed using ISOMAP [38]. Only the inter-sequence distance matrix is required as input, so it was not necessary to explicitly represent each activity sequence as a point in this space.

A plot of the two-dimensional embedding of the 32 activity sequences is shown in Fig. 6. The most striking observation about this plot is that, aside from one outlier, the points are well clustered, and even appear to be linearly separable in this particular embedding. This would indicate that a large majority of the variation between these reconstructed activity sequences can be explained by only two degrees of freedom. This also explains why we were able to achieve such a high classification accuracy (see Table 1). Most importantly, this illustrates that the reconstruction of only a single point on the body can contain a significant amount of information about the motion of the subject.

10. Application: clinical gait analysis

In the following set of experiments we wish to explore the applicability of the proposed reconstruction techniques to clinical gait analysis. Gait is an important cue regarding a person’s health, and often contains a wealth of information regarding the existence
and severity of certain medical conditions [25]. Traditionally, gait analysis is performed manually by an expert physician who simply observes the gait of a patient in order to assess the degree of his/her debilitation. However, results of such a subjective manual analysis may vary from one physician to another for the same patient. In light of the current state-of-the-practice, it may be both possible and desirable to standardize this type of gait analysis through the use of technology. This may also make the results more accurate, precise, and repeatable.

One possibility is to use commercial motion capture systems in order to collect accurate data about gait (see, for example, [28], among others). However, such systems are composed of multiple infra-red cameras (a minimum of six), are quite expensive, and require some technical expertise to operate, making them impractical in many clinical settings. More recently, there have also been attempts to introduce further automation into the diagnostic procedure in the form of single-camera systems [26,27]. These works are more preliminary, and being single-camera systems which perform image-based analysis, suffer from the same drawbacks of view-invariance as more general single-view techniques.

Using the techniques proposed in this work for the 3D reconstruction of periodic motion, we aim here to demonstrate that it is possible to perform clinical gait analysis using only a single camera, in a way that is view-invariant. Specifically, we wish to show that these reconstructions contain rich enough information to be used in such a demanding application area. Note that the analysis performed here is only exploratory in nature, and is intended as a proof-of-concept — clearly further work would need to be done before these results can be applied in a clinical setting. Furthermore, we note that the use of marker-based tracking (used here for some of the data collection) is an acceptable constraint in a clinical setting.

In this set of experiments we explore the specific case of people whose gaits have been affected by strokes, and we compare them to the gaits of healthy people. In particular, we concern ourselves here with patients who exhibit paretic gait, in which one side of the body has been partially paralyzed by a stroke. We will show that it is possible to distinguish between the gaits of these patients and those of healthy individuals. Furthermore, our results indicate that it may be possible to infer the degree of debilitation exhibited from only a small number of points on the body.

10.1. Data collection

The data used for these experiments consisted of gait information from both healthy people and people who have suffered from strokes. In all cases, the points on the body that were tracked included the subjects' ankle and knee joints — this is the only information that was used from each subject. The data was collected from the natural gaits of these subjects, in that they were asked only to walk in a straight line, without any further instruction or aid.

10.1.1. Healthy gait

The set of healthy gait data consisted of samples from six different people without any known medical conditions which affect their movement. Several sequences were filmed of each person, from both left- and right-side viewing angles, using a standard resolution video camcorder. Small colored markers were attached to the ankle and knee joints over the subjects' clothing, and they were tracked using color classification. In total, there were 27 left-side and 32 right-side track sequences of healthy gait.

10.1.2. Paretic gait

Samples of debilitated gait came from eight different individuals exhibiting partial paralysis as a result of a stroke (3 left-side paretic and 5 right-side paretic). Several sequences were collected from each subject using a commercial motion capture system, which provides accurate 3D trajectories of infra-red markers attached to the subject's body. Since many of the sample trajectories consisted of just one full stride of gait, longer sequences were synthesized by concatenating three periods of motion from the raw data. Then, two virtual camera positions and orientations were chosen (such that the left and right legs, respectively, could be seen), and the 3D trajectories were projected into the image coordinates of these virtual cameras. This comprised the data which we used as input for these experiments. Note that this is similar to the tracking approach taken in related work such as [5], except that here we use the more realistic perspective projection instead of the orthographic assumption. As before, only tracks of the subjects' ankle and knee joints were used. In total, there were 7 tracks of left-side paretic gait, and 13 of right-side paretic gait.

10.2. Trajectory reconstruction

Given the image-coordinate tracks described in the previous section, the constant period of motion was estimated using the technique described in Section 3.3. Since the period of motion is a property of the body as a whole (as opposed to being different for each body part), only the estimate from the ankle joint of each sequence was used. Next, each image-coordinate trajectory was resampled using Gaussian Process regression to get 100 samples per period. These resampled image-coordinate trajectories were then reconstructed using the 3D geometric error cost function (see Section 3.2), and the reconstruction was performed using the constant period formulation. Finally, since each sequence in these gait samples consisted of tracks from two rigidly-connected body points (i.e., ankle and knee), each reconstructed knee-joint trajectory was scaled relative to the corresponding reconstructed ankle trajectory in order to enforce the rigidity constraint (Section 4).

10.3. Gait distance computations

Before any high-level analysis can be performed, we must first compute the distances between all pairs of reconstructed ankle/knee sequences. We have already enforced the rigid body constraints between the knee and ankle for each sequence. However, we must also appropriately scale each entire ankle/knee pair (in 3D) in a way that enables us to compute scale-invariant distances. This is accomplished by choosing one period of one ankle trajectory as a reference period. Then, for each sequence, every one-period segment of ankle trajectory is scaled, rotated, and translated so that it is best-aligned with the reference period (using the OPD, see Section 6). For the one-period segment that most closely matches the reference period, the scale-factor computed for that alignment is then used as the overall scaling of that sequence.

Once each sequence has been appropriately scaled, we can compute the distance between each pair of reconstructed ankle/knee trajectories (i.e., the inter-sequence distances). Distances are computed using the OPD, which calculates the translation and rotation that minimizes the MSE between the two sets of points, including both the ankle and knee reconstructions.

10.4. Sequence classification

In this particular experiment we begin to examine our ability to classify between healthy and paretic gait using the distances computed above. Here we consider each ankle/knee gait sequence independently, ignoring (for now) our knowledge of which subject it was generated by. All healthy sequences are used, but only the paretic-side gait of the stroke victims are included in this experiment.
A simple nearest-neighbor classification is used, in which a test sequence is assigned the label of its nearest-neighbor in the training data. We choose 10-fold cross-validation in order to test the classification performance. Note that our goal here is not to develop a sophisticated new classification scheme, but only to explore the informativeness of these reconstructions as a proof-of-concept pilot study. The classification results are summarized in Table 2. Note that overall classification accuracy is close to 95% for individual gait sequences.

### 10.5. Subject classification

So far we have treated each gait sequence individually, regardless of which subject it was from. However, taking the identity of the subject into account may aid in performing a more robust analysis. In this experiment, we consider a more realistic case, which is perhaps more similar to the way analysis would be performed in a clinical setting. Instead of measuring only inter-sequence distances, here we consider inter-person distances, taking into account all available gait sequences for each subject.4 Here the goal is to classify each person or subject as belonging to either the healthy or paretic gait category. Note that this is more similar to the type of automated analysis that could be performed in a doctor’s office: first several gait samples would be collected from the current subject in question, then this person’s gaits would be compared to both the healthy and debilitated subjects in the database in order to make a diagnosis.

To compute the inter-person distance between two subjects, we simply take the median of all inter-sequence distances between these two people, due to its robustness to outliers. In this experiment each subject was classified as either healthy or paretic based on all gait sequences from one side of his/her body. As before, classification was performed here using a simple nearest-neighbor scheme, where in this case a subject was assigned the same label as the subject in the training set with which the minimum inter-person distance was computed. In each case, one subject was used as the test set, while the training set consisted of all remaining subjects (i.e., the database).

Using this classification scheme, we achieved a classification accuracy of 100% for both the healthy and paretic subjects. As can be seen from this result, classification accuracy is improved by treating each subject as a whole, as well as the robustness of the median to outliers. This further supports our assertion that gait reconstructions using only two points on the body can still contain significant information about the subject.

### 10.6. Manifold embedding

We examine the ISOMAP embedding of the gait reconstructions, as before. A plot of the two-dimensional embedding of the healthy and paretic gait sequences is shown in Fig. 7, where healthy gaits are indicated by blue points, and paretic gaits by red points.5 One fact that is quite striking about this embedding is that, barring three outlier points, the healthy and paretic gaits are linearly separable in this low-dimensional space. We can draw two interesting conclusions from this observation. First, this result seems to indicate that most of the variation in this set of data can be explained by only two degrees of freedom. So even though the gaits represented by these high-dimensional points are quite complex motions, most of the difference between them can be parameterized with two variables. Second, this embedding explains our previous results for both the sequence classification and the subject classification tasks. It is clear that, at least in this particular embedding, it would be difficult to develop any classifier which could achieve perfect classification accuracy based only on the individual sequences. This is evidenced by the few outliers depicted in Fig. 7 – the three healthy gaits embedded in the midst of all the paretic data points. However, when we switch to subject classification using inter-person distances, this problem is alleviated, since we take into account all gait sequences from each subject, and small numbers of outliers no longer affect the classification.

Interestingly, these results also indicate that, not only does the manifold embedding indicate linear separability in 2D, but it may also contain information about the degree or severity of the subject’s paralysis. Informally, we can observe that paretic gaits closer to the healthy gaits in the manifold embedding (i.e., red points farther to the right in Fig. 7) show less paralysis than those farther from the healthy gaits. Unfortunately, it is difficult to illustrate the differences in gait in a static image. As such, video clips showing examples of three different gait sequences with different levels of paralysis can be found online at http://www.ece.umn.edu/users/ribn0003/stroredata/. Included are paretic gait sequences from three different parts of the manifold – moving from right to left in Fig. 7. Each sequence shows the motion-capture tracks of the subject’s knees and ankles, where each sequence was synthesized from the real motion-capture as described earlier.

Finally, we emphasize again that the information used in this experiment only consisted of tracks of two points on each subject’s body. This is in contrast to existing work (e.g., [26,27]), which typically requires knowledge about the motion of a much larger number of points on the body. The results presented above, although preliminary, seem to suggest that a significant amount of information is contained in the motion of only a very small number of points on the body, and that techniques for properly extracting and processing such information are necessary.

### 11. Application: subject identification

In the previous sections we demonstrated the usefulness of the proposed method for activity classification as well as clinical gait analysis. Gait is also an important biometric cue for person identification. In this section, we evaluate the applicability of the method to classify subjects performing the same action, i.e., walking. For this experiment, we use motion capture data from the CMU Graphics Lab Motion Capture Database6 and the Georgia Tech HumanID Gait Database [39].

From the CMU database, we selected a subset of 56 walking sequences from five different subjects, which conformed to regular motion along a single direction, with two or more periods in the sequence. The 3D trajectories of left and right knees, ankles, heels and toes from the motion capture data were projected onto two cameras views on opposite sides of the subject. The image tracks were then resampled and used to reconstruct each sequence following the procedure explained in Section 10.2. The rigidity constraints are ensured between the knee–ankle, heel–ankle and toe–ankle connections, with the ankle as the central point. For each

---

3 We take the left- and right-side gaits of healthy individuals separately, and use only the paretic side of debilitated subjects.

4 For interpretation of color in Figs. 1, 2, 4–7, the reader is referred to the web version of this article.

5 http://mocap.cs.cmu.edu/
point on the body, the distances between the different reconstructed trajectories were computed as in Section 10.3. We now have, for each subject and for each camera view, the 3D reconstruction of the six joint-pairs, three on each leg. Following the pattern of the activity classification experiment, each sequence is classified as belonging to a particular subject using an opposite-view nearest-neighbor classification scheme, to demonstrate the view-invariant nature of these reconstructions. Different combinations of the six joint-pairs were tested for classification, and the results are shown in Table 3.

Note here that the classification accuracy is over five subject classes all performing the same action, and their gait variation alone helps in distinguishing the different people. Compare this with the performance of a random classifier on $K = 5$-class classification problem, which would give an accuracy of $1/K = 20\%$. Even with using a single joint-pair, the classification accuracy is noteworthy, and as more joints are used together, the accuracy improves even further.

From the Georgia Tech HumanID dataset, the motion capture data from the entire set of 18 subjects with six sequences each (except subject 7, with only four sequences) was used to reconstruct the trajectories in the same way as with the CMU data. Due to the shorter duration of the sequences, only the left leg had at least two complete periods, and therefore only the three points (foot, lower leg and upper leg) on the left leg were used for our reconstruction. An opposite-view nearest-neighbor classification was performed across two opposing camera views, and the accuracies are shown in Table 4 for using the foot/lower-leg joint, the lower-leg/upper-leg joint, and the two combined. Similar to the previous experiment, using more points on the body resulted in greater accuracy. It is important to note that the accuracies reported here are for a dataset of $K = 18$ classes, and contrast this with the performance of a random classifier, which would have an accuracy of $1/K = 5.56\%$.

In both these experiments, the invariance of the reconstruction across the camera views is demonstrated by classification of different subjects performing the same activity. This experiment can be considered analogous to, for example, using gait for recognition of people across camera views at an airport. It thus could be used to keep track of the different people as they move from one camera’s view to another.

### Table 3

<table>
<thead>
<tr>
<th>Joint-pairs</th>
<th>Legs (%)</th>
<th>Left</th>
<th>Right</th>
<th>Both</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heel–ankle</td>
<td></td>
<td>71.43</td>
<td>76.79</td>
<td>83.93</td>
</tr>
<tr>
<td>Knee–ankle</td>
<td></td>
<td>74.11</td>
<td>83.04</td>
<td>85.71</td>
</tr>
<tr>
<td>Toe–ankle</td>
<td></td>
<td>75.00</td>
<td>76.79</td>
<td>85.71</td>
</tr>
<tr>
<td>Heel–ankle–knee</td>
<td></td>
<td>76.79</td>
<td>82.14</td>
<td>85.71</td>
</tr>
<tr>
<td>Heel–ankle–toe</td>
<td></td>
<td>75.89</td>
<td>75.89</td>
<td>86.61</td>
</tr>
<tr>
<td>Knee–ankle–toe</td>
<td></td>
<td>83.93</td>
<td>78.57</td>
<td>88.39</td>
</tr>
<tr>
<td>All three</td>
<td></td>
<td>83.04</td>
<td>78.57</td>
<td>88.39</td>
</tr>
</tbody>
</table>

### Table 4

<table>
<thead>
<tr>
<th>Joint-pairs</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foot/lower-leg</td>
<td>54.72</td>
</tr>
<tr>
<td>Lower-leg/upper-leg</td>
<td>50.00</td>
</tr>
<tr>
<td>Both</td>
<td>60.38</td>
</tr>
</tbody>
</table>

12. Conclusions and future work

In this paper we have built on our previous work on reconstructing periodic motions in 3D in order to explore its effectiveness in human motion analysis. To this end, we have addressed several of the challenges arising from the complexities inherent in real human motion. This includes the development of a technique for linking reconstructions of multiple point trajectories on an articulated body, an algorithm for iteratively reconstructing cyclic motions whose period is non-constant, and a proposed distance function for measuring the difference between two reconstructions in a robust and consistent manner.

In order to justify the use of these reconstruction techniques, we have also performed an empirical study regarding the periodicity of real human motion. Using motion-capture data from several instances of normal gait, measures of both temporal and spatial periodicity were considered, as well as the suitability of the proposed motion model for periodic trajectories. We have found that human motion can indeed adhere closely to strict periodicity over a duration of 3–4 strides.

Finally, the applicability of the proposed algorithms to the domain of human motion analysis was demonstrated through three proof-of-concept experiments. In the first, the trajectory of one point on the body was reconstructed in order to distinguish between four different activities, independent of the viewing angle. Classification accuracy was 96.8%, with 31 of 32 sequences classified correctly.

In the second study, we demonstrated that reconstructions of two points on the body could be used to accurately classify gait sequences as belonging to either healthy individuals, or to subjects whose movement has been affected by a stroke. A classification accuracy of 100% was achieved when classifying between sets of trajectories consisting of multiple instances of each subject’s gait. Furthermore, it was observed that a low-dimensional manifold embedding of the reconstructed trajectories contained information regarding the severity of the debilitation.

In the third and most challenging task, we attempted to classify different people performing the same gait, using the trajectories of various points on the subject’s legs. Accuracies of 88.4% over a 5-class data set (CMU), and 60.4% over an 18-class data set (Georgia Tech) demonstrate the potential of gait as a biometric cue.

There are several directions that could be pursued in future research along these lines. One possibility is to consider more generalized notions of periodicity, and develop techniques for reconstructing these trajectories. It might also be interesting to explore other application areas for the proposed algorithms,
including possibly 3D pose estimation. In this case it might be necessary to reconstruct the trajectories of a larger number of points on the body.

Acknowledgments

The authors wish to thank Professors David Nuckley and James Carey from the Department of Physical Therapy at the University of Minnesota for providing access to the stroke patient gait data used in this paper. In addition, this material is based upon work supported in part by the National Science Foundation through Grants #IIP-0443945, #CNS-0821474, #IIP-0934327, #CNS-1039741, and #SMA-1028076. The data for subject identification experiments were obtained from CMU Graphics Lab Motion Capture Database (http://mocap.cs.cmu.edu/), funded by NSF EIA-0196217, and the Georgia Tech 'Human Identification at a Distance' Database, funded by the DARPA HID Program.

References


