Technical Report

Department of Computer Science
and Engineering
University of Minnesota
4-192 EECS Building
200 Union Street SE
Minneapolis, MN 55455-0159 USA

TR 09-025

3D Reconstruction of Periodic Motion from a Single View

Evan Ribnick and Nikos Papanikolopoulos

September 24, 2009

3D Reconstruction of Periodic Motion from a Single View

Evan Ribnick and Nikolaos Papanikolopoulos
{ribnick,npapas}@cs.umn.edu

Abstract

Periodicity has been recognized as an important cue for tasks like activity recognition and gait analysis. However, most existing techniques analyze periodic motions only in image coordinates, making them very dependent on the viewing angle. In this paper we show that it is possible to reconstruct a periodic trajectory in 3D given only its appearance in image coordinates from a single camera view. We draw a strong analogy between this problem and that of reconstructing an object from multiple views, which allows us to rely on well-known theoretical results from the multi-view geometry domain and obtain significant guarantees regarding the solvability of the estimation problem. We present two different formulations of the problem, along with techniques for performing the reconstruction in both cases, and an algorithm for estimating the period of motion from its image-coordinate trajectory. Experimental results demonstrate the feasibility of the proposed techniques.

1 Introduction

Periodic motion is quite common in our everyday experience, and one of the most frequent and interesting examples arises from natural human motions such as walking and running. It has been recognized as an important cue in the literature by researchers interested in various areas, such as activity recognition and gait analysis, among others. However, since monocular systems are far more commonplace than multi-camera observations, existing techniques for analyzing periodicities are largely image-based. This typically implies a lack of view-invariance, since the same motion can have a drastically different appearance in the image if viewed from a different angle.

For example, consider the case illustrated in Figure 1. The observed motion is that of a person walking, where a point near his ankle has been tracked. In this case we have two simultaneous views of the same motion, taken from different viewing angles. Notice that even though both views observe the same trajectory in world coordinates, its projection into image coordinates varies significantly as the viewing angle is altered.

Because they are not view-invariant, motion/action classifiers or mappings that are learned solely from image coordinate observations typically do not generalize well to different cameras or viewing angles. In general, a new classifier or mapping must be learned separately for every viewing angle (e.g., Elgammal and Lee (2004)). As such, one way to circumvent this need is to first infer characteristics of the motion in 3D, and then any subsequent analysis is performed in world coordinates, independent of the viewing angle from which the motion was originally imaged. This allows one to develop general classifiers or mappings, which theoretically can be used for any captured sequence. Of course, in the case of any arbitrary motion it may not be possible to reconstruct the object's trajectory in 3D given only its appearance in image coordinates. However, when more is known about the motion of the object in the world, this additional information can sometimes be used to adequately constrain the reconstruction problem (see, for example, Ribnick et al. (2009)).

In this paper we explore the idea of reconstructing periodic point trajectories in 3D given only their appearances in image coordinates. Broadly speaking, the goals of this paper are as follows: to show that such reconstructions are both possible to obtain and accurate in realistic settings, and to provide insight regarding the solvability of this reconstruction problem under different geometries. The foundation of this work is the idea that periodicity provides a physical constraint on the trajectory (i.e., a physical motion model), making it possible to infer the path of motion in world coordinates in most cases. We present two possible formulations of this reconstruction problem, resulting in two different cost functions. Suitable optimization strategies are proposed for each of them, and their sensitivities to noise and ability to reconstruct from real data are analyzed. A novel

Figure 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.

technique for estimating the period of motion from an image-coordinate trajectory, in which the periodicity of the original signal may be distorted, is also developed.

Importantly, we perform a thorough theoretical analysis of this reconstruction problem, which yields significant guarantees and insight regarding solvability under different imaging geometries. This analysis is based on an analogy, in which we rely on the fact that the problem at hand can be shown to be mathematically equivalent to reconstructing a single object from multiple views under certain conditions. This allows us to draw on the rich body of existing theory in the domain of multi-view geometry and 3D reconstruction.

The focus of this work is not on the specifics of tracking points of interest – instead, we choose to focus on the problem of reconstructing periodic 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. A similar assumption was made in the related work Zhang and Troje (2007) (in fact, we operate under fewer assumptions here). In any case, human body tracking is an active area of research in itself – see, for example, (Fossati et al., 2008; Ramanan et al., 2007; Sigal and Black, 2006) – and therefore we can assume that it is only a matter of time before more reliable and universally applicable solutions are developed for this particular problem.

The rest of this paper is structured as follows. After discussing some related work in §2, §3 introduces some basic definitions and notation. Two different formulations of the reconstruction problem are developed in §4 and §5, and theoretical results regarding solvability are presented in §6. §7 introduces a technique for estimating the period of motion based only on the appearance of a trajectory in the image plane. The feasibility of the proposed methodologies is explored in §8 via thorough experimentation. Finally, §9 presents some conclusions and possibilities for future work.

2 Related Work

This work is unique in that it is the first (to the best of the authors' knowledge) to propose techniques for reconstructing periodic motions without requiring training data, and as such is fully generalizable to many types of periodic motion. Some parts of this work appeared previously in the conference papers Ribnick and Papanikolopoulos (2008, 2009). Otherwise, the previous research most closely related to this is that of Zhang and Troje (2007), in which known 3D trajectories are used to learn Fourier coefficient-based representations of common human motions. Their work is primarily concerned with human gait, and is based on the assumptions that (i) observed motions are similar to those in the training data, (ii) images are formed via orthographic projection, and (iii) image tracks of points of interest are available.

Our approach is fundamentally different than this previous work because we infer the 3D trajectory directly using only geometric constraints. Since no training data is required, these algorithms are fully generalizable to any type of periodic motion, opening

the door to a whole new set of application areas in which it does not matter if the observed type of periodic motion has been seen before. Finally, since our work takes the full perspective projection model into account, it allows for motions which contain a translational component (i.e., are non-stationary), increasing its applicability in real-world environments.

Belongie and Wills (2004) propose a technique for estimating the structure of an object that is undergoing periodic motion. They consider snapshots of the object separated by one period in time, and treat them as multiple views of the same object. These multiple views are used to perform geometric inference using techniques from multi-view geometry. Note that this is fundamentally different than what we aim to achieve in this paper, since our goal is to estimate the trajectory of an object in 3D, and not the structure of the object itself. As such, in their work it is necessary to make use only of one image per period of motion, while here we estimate the 3D position of the point of interest in every frame.

Other work related to periodic motion has focused mainly on detection and analysis in image coordinates, and does not explicitly consider the 3D information that is embedded in the periodicity. In some cases, the object of interest is first tracked, and translational motion removed before processing. Several techniques (e.g., Polana and Nelson (1997); Liu and Picard (1998); Cutler and Davis (2000); Ormoneit et al. (2005); Orriols and Binefa (2003); Briassouli and Ahuja (2007)) use some type of Fourier analysis of pixel intensities or appearances to detect, segment, and represent periodic motions. Frequency domain techniques have also been used in the activity classification and recognition tasks (Tsai et al., 1994; Little and Boyd, 1995; Fujiyoshi et al., 2004).

Some of the other existing work does not rely on Fourier techniques to detect and analyze periodic motion. Seitz and Dyer (1997) present a framework for analyzing cyclic motions that deviate from pure periodicity. In Allmen and Dyer (1990), cyclic motions are detected as repetitions on surfaces carved out by a moving object in xyt space. Periodicity has also been used to recognize and classify human gestures (Cohen et al., 1996), human facial expressions, and bird movements (Davis et al., 2000).

3 Definitions and Basic Equations

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

$$\mathbf{v}(t+nT) = \mathbf{v}(t), \tag{1}$$

for any integer n, where $\mathbf{v} \triangleq (\dot{X}, \dot{Y}, \dot{Z})$ and T is the period. Notice that this definition differs from that in most of the previous work (see §2), where periodicity was defined in terms of position rather than velocity. In general, motion that is periodic in velocity (as defined above) includes as a special case motion that is periodic only in position. This definition of periodicity also includes motions for which there is translation from one period to the next, such as the foot of a walking person. In fact, we will see later that the translational component is necessary in order to solve the reconstruction problem that is the focus of this paper, which is inherently related to the change in the appearance of the periodic trajectory in the image as the object displaces relative to the camera. This point will be illustrated in detail in §6.

In terms of the 3D position of the point, we have:

$$\mathbf{p}(t+nT) = \mathbf{p}(t) + n\mathbf{\Delta}_{\mathbf{p}_T}, \tag{2}$$

where $\mathbf{\Delta}_{\mathbf{p}_T} \triangleq (\Delta_{X_T}, \Delta_{Y_T}, \Delta_{Z_T})$ is the displacement per period of the point (or equivalently, the inter-period displacement), 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 $\|\mathbf{\Delta}_{\mathbf{p}_T}\|_2$.

We next consider the displacement between two samples from the same period. For some length of time $\tau < T$:

$$\mathbf{p}(t+\tau) = \mathbf{p}(t) + \delta_{\mathbf{p}}, \tag{3}$$

where $\delta_{\mathbf{p}} \triangleq (\delta_X, \delta_Y, \delta_Z)$ is the 3D displacement between the samples at times t and $t+\tau$. Note that this displacement is the same for any pair of samples taken at times $t+nT$ and $t+nT+\tau$, for any integer n, but is different for each value of τ.

Since samples are taken at discrete times determined by the video frame rate, we represent times using discrete indices of the form t_k^i. This represents the time of the k-th sample in the i-th period, where $k = 0, 1, ..., N-1$ and $i = 0, 1, ..., M-1$, and N and M are the number of samples per period and number of periods, respectively. We can then arrive at the

following expression for the position at time t_k^i:

$$\mathbf{p}_k^i = \mathbf{p}_0^0 + i\boldsymbol{\Delta}_{\mathbf{p}_T} + \delta_{\mathbf{p}_k}, \tag{4}$$

which is written in expanded form as:

$$\begin{pmatrix} X_k^i \\ Y_k^i \\ Z_k^i \end{pmatrix} = \begin{pmatrix} X_0^0 \\ Y_0^0 \\ Z_0^0 \end{pmatrix} + i \begin{pmatrix} \Delta_{X_T} \\ \Delta_{Y_T} \\ \Delta_{Z_T} \end{pmatrix} + \begin{pmatrix} \delta_{X_k} \\ \delta_{Y_k} \\ \delta_{Z_k} \end{pmatrix}. \tag{5}$$

This equation expresses the sample at time t_k^i in terms of the zeroth sample in the zeroth period, t_0^0. In some cases we will instead write this as:

$$\begin{pmatrix} X_k^i \\ Y_k^i \\ Z_k^i \end{pmatrix} = \begin{pmatrix} X_k^0 \\ Y_k^0 \\ Z_k^0 \end{pmatrix} + i \begin{pmatrix} \Delta_{X_T} \\ \Delta_{Y_T} \\ \Delta_{Z_T} \end{pmatrix}, \tag{6}$$

when the formulation does not depend on $\delta_{\mathbf{p}_k}$ – in this case the sample at time t_k^i is expressed relative to t_k^0, the k-th sample in the zeroth period.

In order to see how samples from the periodic trajectory are projected into image coordinates, we rely on the pinhole camera model to obtain equations for the image point (u_k^i, v_k^i) in pixel coordinates:

$$\begin{pmatrix} u_k^i \\ v_k^i \end{pmatrix} = \frac{1}{Z_k^i} \begin{pmatrix} f_x & 0 \\ 0 & f_y \end{pmatrix} \begin{pmatrix} X_k^i \\ Y_k^i \end{pmatrix} + \begin{pmatrix} c_x \\ c_y \end{pmatrix}. \tag{7}$$

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 pixel units.

Equation (7) can be expanded using (5) as follows:

$$\begin{aligned} \begin{pmatrix} u_k^i \\ v_k^i \end{pmatrix} = & \frac{1}{Z_0^0 + i\Delta_{Z_T} + \delta_{Z_k}} \\ & \times \begin{pmatrix} f_x & 0 \\ 0 & f_y \end{pmatrix} \begin{pmatrix} X_0^0 + i\Delta_{X_T} + \delta_{X_k} \\ Y_0^0 + i\Delta_{Y_T} + \delta_{Y_k} \end{pmatrix} + \begin{pmatrix} c_x \\ c_y \end{pmatrix}, \end{aligned} \tag{8}$$

or it can be expanded using (6):

$$\begin{pmatrix} u_k^i \\ v_k^i \end{pmatrix} = \frac{1}{Z_k^0 + i\Delta_{Z_T}} \begin{pmatrix} f_x & 0 \\ 0 & f_y \end{pmatrix} \begin{pmatrix} X_k^0 + i\Delta_{X_T} \\ Y_k^0 + i\Delta_{Y_T} \end{pmatrix} + \begin{pmatrix} c_x \\ c_y \end{pmatrix}. \tag{9}$$

Thus, we see that it is possible to express the projection of any sample projected into image coordinates, (u_k^i, u_k^i), as a function of the sample at time t_0^0, the inter-period displacement $\boldsymbol{\Delta}_{\mathbf{p}_T}$, and the intra-period displacement $\delta_{\mathbf{p}_k}$ (8), or as a function of the corresponding sample at time t_k^0 and the inter-period displacement $\boldsymbol{\Delta}_{\mathbf{p}_T}$ (9).

4 Minimizing the Reprojection Error

This section introduces the first of two different formulations for the 3D reconstruction of a periodic trajectory from a single camera view. As is customary in this type of geometric inference problem, the reconstruction is posed as an optimization problem, in which the optimal solution is obtained by minimizing a cost function. In this case, the cost function is comprised of the sum of reprojection errors:

$$F(\mathbf{X}_1) \triangleq \sum_{i=0}^{M-1} \sum_{k=0}^{N-1} \left\| \begin{pmatrix} u_k^i \\ v_k^i \end{pmatrix} - \begin{pmatrix} \hat{u}_k^i \\ \hat{v}_k^i \end{pmatrix} \right\|_2^2, \tag{10}$$

where M is the number of periods, and N is the number of samples in each period. The reprojection error consists of the Euclidean distance between the observed image coordinate track, (u_k^i, v_k^i), and the reprojection of the trajectory into image coordinates based on the current estimate of the reconstruction, $(\hat{u}_k^i, \hat{v}_k^i)$. The reprojections are formed based on the projection given by Equation (8).

The reconstruction here is parameterized by the vector $\mathbf{X}_1$, which contains the very first sample (X_0^0, Y_0^0, Z_0^0), the inter-period displacement $(\Delta_{X_T}, \Delta_{Y_T}, \Delta_{Z_T})$, and the displacements within each period $(\delta_{X_k}, \delta_{Y_k}, \delta_{Z_k})$, $k = 1, 2, ..., N-1$. This results in $6 + 3(N-1)$ variables. Note that it is only possible to obtain reconstructions that are accurate up to a scale factor. As such, in practice the reconstruction is performed by fixing one of the variables to an arbitrary value and minimizing over the remaining variables, resulting effectively in $\mathbf{X}_1 \in \mathbb{R}^{5+3(N-1)}$.

The variables (X_0^0, Y_0^0, Z_0^0) and $(\delta_{X_k}, \delta_{Y_k}, \delta_{Z_k})$, $k = 1, 2, ..., N-1$ parameterize the first period of motion. Since all periods are merely shifted copies of the first, all other periods can then be formed using the inter-period displacement $(\Delta_{X_T}, \Delta_{Y_T}, \Delta_{Z_T})$. Together, these $6+3(N-1)$ variables fully characterize the periodic trajectory in 3D. The optimal estimate is then given by:

$$\mathbf{X}_1^* = \arg\min F(\mathbf{X}_1). \tag{11}$$

4.1 Local Convexity

The cost function $F(\mathbf{X}_1)$ can be rewritten as a sum of ratios of quadratic functions of the form:

$$F(\mathbf{X}_1) = \sum_{i=0}^{M-1} \sum_{k=0}^{N-1} \frac{\mathbf{X}_1^T \mathbf{A}_k^i \mathbf{X}_1 + (\mathbf{B}_k^i)^T \mathbf{X}_1 + C_u^i}{\mathbf{X}_1^T \mathbf{D}_k^i \mathbf{X}_1 + (\mathbf{E}_k^i)^T \mathbf{X}_1 + F_k^i}, \tag{12}$$

where $\mathbf{A}_k^i$ and $\mathbf{D}_k^i$ are square coefficient matrices, $\mathbf{B}_k^i$ and $\mathbf{E}_k^i$ are vectors, and $\mathbf{C}_k^i$ and $\mathbf{F}_k^i$ are scalars. As such, $F(\mathbf{X}_1)$ is not globally convex in general — even a sum of linear-fractional functions is known to be nonconvex, and cannot be solved efficiently using global methods for more than ten ratios (Schaible and Shi, 2003). Furthermore, the individual subfunctions in the summation are nonconvex themselves, since they are ratios of quadratic functions. However, we have found that, in practice, $F(\mathbf{X}_1)$ is locally convex around its optimal solution, and that this convex region is typically large, even in the presence of significant measurement noise.

Given an initial solution, a local optimization algorithm can then be used to minimize the cost function. In this case we use the Levenberg-Marquardt (LM) algorithm (Nocedal and Wright, 1999). LM is convenient since it automatically interpolates between quasi-Newton's method (more efficient but less stable) in more convex regions, and gradient descent (less efficient but more stable) in less convex regions.

4.2 Initial Solution

Since the cost function (10) typically contains a large convex region around the optimum but is nonconvex otherwise, it is extremely important to obtain an initial solution which is inside this region to avoid converging to one of the local minima. We obtain an initial solution by splitting the problem into two subproblems, and obtaining separately an analytic solution for each one, according to the procedure described in (Ribnick and Papanikolopoulos, 2008).

5 Minimizing the 3D Geometric Error

This section introduces another formulation. In this case the cost function is formed by minimizing a 3D geometric error. As will be demonstrated, it has some significant advantages over the previous cost function (§4).

Rearranging the terms in Equation (9), we can obtain expressions for X_k^0 and Y_k^0 in terms of estimates of Z_k^0 and $(\Delta_{X_T}, \Delta_{Y_T}, \Delta_{Z_T})$:

$${}^i\hat{X}_k^0 = \frac{u_k^i - c_x}{f_x}(\hat{Z}_k^0 + i\hat{\Delta}_{Z_T}) - i\hat{\Delta}_{X_T} \tag{13}$$

$${}^i\hat{Y}_k^0 = \frac{v_k^i - c_y}{f_y}(\hat{Z}_k^0 + i\hat{\Delta}_{Z_T}) - i\hat{\Delta}_{Y_T}, \tag{14}$$

where " $\hat{}$ " denotes that a quantity is an estimate, and ${}^i\hat{X}_k^0$ and ${}^i\hat{Y}_k^0$ 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 $k = 0, 1, ..., N-1$ and each period $i = 0, 1, ..., M-1$. Note that, from these relations, it is clear that the parameters we wish to estimate, Z_k^0, $k = 0, 1, ...N-1$ and $(\Delta_{X_T}, \Delta_{Y_T}, \Delta_{Z_T})$, completely characterize the trajectory of the point in 3D.

Ideally ${}^{i_1}\hat{X}_k^0 = {}^{i_2}\hat{X}_k^0$ and ${}^{i_1}\hat{Y}_k^0 = {}^{i_2}\hat{Y}_k^0$ for any sample k and any pair of periods i_1 and i_2. Therefore, making use of (13) and (14), we can obtain a pair of equations as follows:

$$\begin{aligned} {}^{i_1}\hat{X}_k^0 - {}^{i_2}\hat{X}_k^0 = \frac{u_k^{i_1} - u_k^{i_2}}{f_x}\hat{Z}_k^0 + (i_2 - i_1)\hat{\Delta}_{X_T} \\ + \frac{i_1 u_k^{i_1} - i_2 u_k^{i_2} + c_x(i_2 - i_1)}{f_x}\hat{\Delta}_{Z_T} = 0 \end{aligned} \tag{15}$$

and

$$\begin{aligned} {}^{i_1}\hat{Y}_k^0 - {}^{i_2}\hat{Y}_k^0 = \frac{v_k^{i_1} - v_k^{i_2}}{f_y}\hat{Z}_k^0 + (i_2 - i_1)\hat{\Delta}_{Y_T} \\ + \frac{i_1 v_k^{i_1} - i_2 v_k^{i_2} + c_y(i_2 - i_1)}{f_y}\hat{\Delta}_{Z_T} = 0. \end{aligned} \tag{16}$$

Two equations of the form (15) and (16) can be obtained for every sample k, for every pair of periods i_1 and i_2. This results in a total of $2N\binom{M}{2}$, 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 the overconstrained homogeneous linear system:

$$A\mathbf{X}_2 = \mathbf{0}, \tag{17}$$

where $A \in \mathbb{R}^{\left(2N\binom{M}{2}\right)\times(N+3)}$ is the coefficient matrix, and $\mathbf{X}_2 \in \mathbb{R}^{N+3}$ is the vector of the parameters

we wish to estimate:

$$\mathbf{X}_2 \triangleq \begin{pmatrix} \hat{Z}_0^0 & \hat{Z}_1^0 & \hat{Z}_2^0 & \dots & \hat{Z}_{N-1}^0 & \hat{\Delta}_{X_T} & \hat{\Delta}_{Y_T} & \hat{\Delta}_{Z_T} \end{pmatrix}^T \tag{18}$$

Since the system (17) is overconstrained, the estimation is cast as a homogeneous linear least-squares problem in which we aim to solve the following optimization problem:

$$\begin{aligned} &\text{minimize } \|A\mathbf{X}_2\|_2 \\ &\text{subject to } \|\mathbf{X}_2\|_2 = 1. \end{aligned} \tag{19}$$

Note that ideally the solution $\mathbf{X}$ must lie in the nullspace of A, and that the nullity of A is one. This implies that it is possible to obtain a solution only up to a scaling factor, which is sufficient for most applications.

Several important points should be made about this new cost function. First, the minimization (19) can be performed efficiently using one Singular Value Decomposition (SVD), where $A = U\Sigma V^T$, and the minimizer $\mathbf{X}_2^*$ is the last column of V (Hartley and Zisserman, 2003). This is a significant advantage over the previous formulation (§4), where an iterative and computationally expensive local optimization algorithm was required to minimize the cost function. Note that, even though numerical solutions for the SVD are internally iterative, very efficient implementations are available since this is a well-studied problem. Second, the cost function $\|A\mathbf{X}_2\|_2$ is convex (as is any linear least-squares cost function), so it is guaranteed that we can arrive at its global minimum every time. Note that this does not provide any guarantees regarding the quality of the solution – the global minimum of the cost function is not necessarily in the same location as the true optimal solution, since the coefficient matrix A may be constructed from noisy image-coordinate samples. Later we will compare the sensitivity to noise of the two cost functions.

Finally, once the parameters $\mathbf{X}_2$ are estimated, the trajectory can be reconstructed using (13) and (14). We estimate X_k^0 by taking the mean of the estimates ${}^i\hat{X}_k^0$ over all periods i, and similarly for Y_k^0, for all samples k. Points from subsequent periods can then easily be reconstructed from the estimates of (X_k^0, Y_k^0, Z_k^0) using the relation (6), $k = 1, 2, ..., N-1$.

5.1 Regularized Reconstruction

The reconstruction technique described above is adequate, but is sensitive to noise in the tracked image-coordinate trajectory. It is possible to add regularization terms to the optimization to enforce smoothness, while still allowing the problem to be solved by a single SVD. The most straightforward way to do this is to add a quadratic smoothing term to the cost function, which seeks to minimize the gradients or the curvatures of the reconstruction (Boyd and Vandenberghe, 2004). Note that these regularization terms are added, not for conditioning (the optimization as described above is already well-conditioned), but only to obtain smoother reconstructions.

The regularized cost function is shown below:

$$\begin{aligned} &\|A\mathbf{X}_2\| + \delta\|D_x\mathbf{X}_2\| + \delta\|D_y\mathbf{X}_2\| + \delta\|D_z\mathbf{X}_2\| \\ &\quad + \gamma\|D_{xx}\mathbf{X}_2\| + \gamma\|D_{yy}\mathbf{X}_2\| + \gamma\|D_{zz}\mathbf{X}_2\|, \end{aligned} \tag{20}$$

where D_x, D_y, and D_z approximate the X, Y, and Z gradients, and D_{xx}, D_{yy}, and D_{zz} approximate the curvatures of the reconstruction, respectively. The scalars δ and γ are weighting coefficients.

6 Multi-View Geometry: An Analogy

An interesting observation is that the problem we are interested in (*3D reconstruction of periodic motion from a single view*) is mathematically equivalent to multi-view reconstruction of a scene. Since it is assumed that each period of the trajectory is identical in world coordinates with only a translation between them, observation of M periods from a single camera is mathematically equivalent to observing a single period from M cameras. Additionally, the analogy implies that the motion between these "virtual cameras" is purely translational, and that the translation from camera i to camera $i+1$ is $(-\Delta_{X_T}, -\Delta_{Y_T}, -\Delta_{Z_T})$. This is illustrated in Figure 2. Recognizing this equivalence allows us to understand our reconstruction problem in a quite intuitive manner, and to draw some powerful conclusions regarding its solvability based on well-known results from multi-view geometry. In the following, we often refer to virtual cameras and the virtual object[1], and use these terms interchangeably with their equivalent entities according to this analogy.

In light of the multi-view analogy, all equations presented thus far can be reinterpreted, taking k to be the index of a point on the virtual object, and i

[1]The term "virtual object" refers to the first period of motion. We imagine that, instead of being samples along a time-series trajectory, these are simply points on some rigid object in 3D.

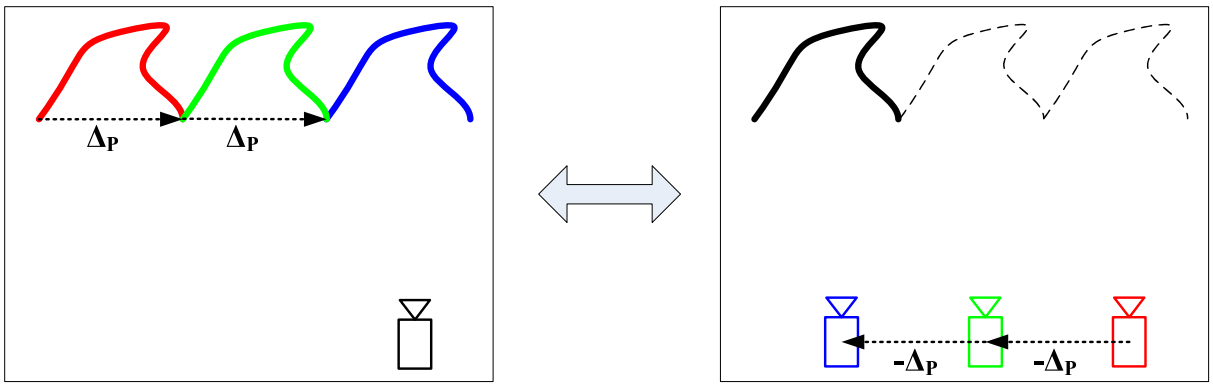

Figure 2: Illustration of the multi-view geometry analogy. Viewing M periods from one camera is mathematically equivalent to viewing 1 period from M cameras.

to be the index of the virtual camera view. For example, recall Equation (9) describing the projection of a periodic motion into image coordinates, repeated here for convenience:

$$\begin{pmatrix} u_k^i \\ v_k^i \end{pmatrix} = \frac{1}{Z_k^0 + i\Delta_{Z_T}} \begin{pmatrix} f_x & 0 \\ 0 & f_y \end{pmatrix} \begin{pmatrix} X_k^0 + i\Delta_{X_T} \\ Y_k^0 + i\Delta_{Y_T} \end{pmatrix} + \begin{pmatrix} c_x \\ c_y \end{pmatrix}. \tag{21}$$

Under this analogy, (21) can reinterpreted as expressing the pixel coordinates of sample k in the i-th camera view, as a function of the 3D position of point k in the coordinate frame fixed at virtual camera $i = 0$, and the translation between the camera views.

Finally, some general remarks about the multi-view interpretation and its applicability. One of the subtleties of this analogy is that the displacements between the virtual cameras must be such that, when the images of the virtual object from all M virtual cameras are overlayed, they must align to form a single contiguous trajectory in image coordinates. This is a result of the original scenario: the 3D trajectory, consisting of M subsequent periods of motion, will project to a smooth trajectory in image coordinates. In addition, it is important to note that this analogy holds equally well for both methods of reconstructing periodic trajectories – minimizing reprojection error (§4) and minimizing 3D geometric error (§5). These are just different techniques for arriving at a solution – the underlying problem is the same no matter which cost function is used.

6.1 Reconstruction from Two Views

It is natural to ask the following question:

- *How many periods of motion are required in order to perform reconstruction?*

The most straightforward way to answer this is to view the problem from the perspective of the multi-view analogy, and rely on established results from that domain. Since the periodic motion problem is equivalent to reconstructing a single virtual object from M camera views, it becomes clear that the minimum number of camera views necessary is two. This is equivalent to the well-known case of binocular stereo. The projective reconstruction theorem (Hartley and Zisserman, 2003) guarantees that, even if the camera's intrinsic parameters were unknown, it would be possible to perform a reconstruction that is accurate up to a projective transformation. Note that this includes not only estimating the 3D locations of the points on the virtual object (triangulation), but also estimating the relative positions of the virtual cameras (calibration).

Translated back to the problem of periodic motion reconstruction, this gives us the following result:

Result 1 *In order to perform 3D reconstruction of periodic motion, the constraint on the number of periods observed is* $M \geq 2$.

Furthermore, in our particular case the camera's intrinsic parameters (f_x, f_y, c_x, c_y) will be known, yielding the even stronger guarantee of a metric reconstruction (i.e., accurate up to a translation, rotation, and scale).

6.2 Number of Point correspondences

Next we attempt to answer the question:

- *How many samples per period are required to perform reconstruction?*

As before, we turn to multi-view geometry for an answer, and rely on established theory. Since the motion between the virtual cameras is purely translational, it turns out that in this case a minimum of only two point correspondences are required in order to perform reconstruction with two camera views (Hartley and Zisserman, 2003). In other words, with only two samples from the virtual object, it is possible to both triangulate their locations and estimate the inter-camera displacement.

In terms of periodic motion reconstruction, this implies the following result:

Result 2 *In order to perform 3D reconstruction of periodic motion, the constraint on the number of samples per period is $N \geq 2$, so long as the number observed periods $M \geq 2$.*

Note that this is the theoretical minimum, so the inter-period displacement $(\Delta_{X_T}, \Delta_{Y_T}, \Delta_{Z_T})$ can be estimated when $N = 2$. However, a reconstruction based on only two samples per period may be very susceptible to noise, and the resolution of any such reconstruction would be too low to be of much use.

6.3 Degeneracy

Finally, we attempt to answer the following question:

- *In which cases is it not possible to perform reconstruction?*

In other words, we wish to uncover geometric configurations of the virtual camera and points on the virtual object in which the reconstruction degenerates. We rely on existing theory, applied to our problem according to the multi-view analogy. We focus here on the limiting case when the number of periods $M = 2$, since according to Result 1 this is the minimum number required, and any additional periods of motion contain more information from which the problem can be solved.

6.3.1 No Translation

First we consider the simple case where the translation between the virtual cameras, $(-\Delta_{X_T}, -\Delta_{Y_T}, -\Delta_{Z_T})$, is zero. Since there is no rotation between them, this means that all M virtual cameras have identical views. Clearly it is not possible to estimate the structure of an arbitrary object from only one view. Therefore, in terms the reconstruction of periodic motion, the following conclusion can be drawn:

Result 3 *In order to perform 3D reconstruction of periodic motion, the inter-period displacement $(\Delta_{X_T}, \Delta_{Y_T}, \Delta_{Z_T})$ must be non-zero.*

Albeit it an obvious degeneracy in terms of multi-view geometry, this has significant implications regarding the applicability of the techniques proposed here. It makes clear that it will not be possible to perform reconstruction in several typical cases of stationary periodic motion, such as a person walking in place on a treadmill, or performing a repetitive gesture, both of which lack a translational component. This also makes sense intuitively in terms of the reconstruction techniques presented, since they rely on differences in appearance in image coordinates between subsequent periods of motion. If there is no difference in appearance, there is not sufficient information from which to draw inferences.

6.3.2 Coplanar Points

It is well-known from multi-view geometry that, in the limiting case when the number of cameras is $M = 2$, it is not possible to perform reconstruction in general when the points on the object are coplanar. However, when the motion between the cameras is purely translational, with no rotational component (as is the case with our virtual cameras, according to the analogy), this is no longer a degeneracy (Hartley and Zisserman, 2003). This is true for any configuration of translated cameras and coplanar world points, except when the world points are also coplanar with the two camera centers.

In terms of our problem with periodic motion, this may be understood as follows. Points on the virtual object correspond to samples from one period of motion in the world. So coplanarity of points on the virtual object only implies that the samples from one period of motion must be coplanar. However, in order for the virtual cameras to also become coplanar with the virtual object, then the entire periodic trajectory must be coplanar in the world, and the camera must also be located on the plane. Thus, we arrive at the result:

Result 4 *It is not possible to perform 3D reconstruction of periodic motion when the number of periods $M = 2$, and the following conditions are satisfied:*

(i) *the trajectory of one period of motion (i.e., the virtual object) lies on a plane in 3D, and*

(ii) *the inter-period displacement $(\Delta_{X_T}, \Delta_{Y_T}, \Delta_{Z_T})$ is coplanar with that trajectory, and*

(iii) *the camera center is also coplanar with that trajectory.*

6.3.3 Other Degeneracies

An additional result states that an object viewed from two cameras cannot be reconstructed from point correspondences if both camera centers and the 3D points lie on a (degenerate or non-degenerate) ruled quadric surface (Hartley and Zisserman, 2003). This may include hyperboloids, cones, cylinders, and planes, plus certain combinations thereof. In these cases, the reconstruction problem also becomes degenerate. However, when passed back through the multi-view analogy and applied to the problem of 3D reconstruction of periodic motion, these results are not useful or informative in realistic situations. Specifically, a periodic trajectory and camera geometry whose analogy (i.e., the corresponding virtual cameras and virtual object) satisfies one of these conditions would have to contain disjoint periods, which would imply extremely large discontinuities in the instantaneous velocity. As such, these degeneracies can be assumed not to exist for all trajectories of interest.

7 Period Estimation

Until now it has been assumed that the period of motion, denoted T, is known. In this section we develop a new technique for estimating the period based only on the appearance of the trajectory in image coordinates. This is a nontrivial problem, since the trajectory, which is assumed to be perfectly periodic (in velocity) in the world, may have been significantly distorted when projected into image coordinates.

7.1 Fourier Analysis in the Image Plane

We consider the case of trajectories which are perfectly periodic in world coordinates according to our definition. One tool for estimating the dominant frequency of motion (the inverse of the period) which is immediately obvious is the Fourier transform. In this subsection we examine the properties of the image trajectory in the frequency domain, and how it relates to the spectrum of the original trajectory in the world. For the sake of simplicity, we make use of the following substitution:

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} f_x & 0 \\ 0 & f_y \end{pmatrix}^{-1} \left\{ \begin{pmatrix} u(t) \\ v(t) \end{pmatrix} - \begin{pmatrix} c_x \\ c_y \end{pmatrix} \right\}, \quad (22)$$

where the image point $(u(t), v(t))$, which is in pixel coordinates, is replaced by the world coordinate equivalent $(x(t), y(t))$. This is convenient, since the image plane then corresponds to the XY-plane of the global coordinate system. In addition, we focus mostly on the image-plane signal $x(t)$, but the same analysis holds for $y(t)$.

7.1.1 Affine Camera

According to Hartley and Zisserman (2003), an affine camera may be assumed when (i) the depth relief (Δ_{Z_T}) is small compared to the average depth, and (ii) the distance of the point from the principal ray is small. In practice, this is a reasonable approximation in many cases.

In an affine camera, the trajectory of a point in image coordinates can be expressed as a function of the trajectory in the world as follows:

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \end{pmatrix} \begin{pmatrix} X(t) \\ Y(t) \\ Z(t) \end{pmatrix} + \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}, \quad (23)$$

where we have used the continuous time index t. Since the motion is periodic in velocity, we write the first derivative of the trajectory:

$$\frac{d}{dt} \left\{ \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} \right\} = \begin{pmatrix} \dot{x}(t) \\ \dot{y}(t) \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \end{pmatrix} \begin{pmatrix} \dot{X}(t) \\ \dot{Y}(t) \\ \dot{Z}(t) \end{pmatrix}. \quad (24)$$

The Fourier transform of the the $\dot{x}(t)$ signal is:

$$\mathscr{F}(\dot{x}(t)) = m_{11}\mathscr{F}(\dot{X}(t)) + m_{12}\mathscr{F}(\dot{Y}(t)) + m_{13}\mathscr{F}(\dot{Z}(t)), \quad (25)$$

and similarly for $\dot{y}(t)$. This is just a linear combination of the Fourier transforms of the velocities in world coordinates. The original motion is periodic in velocity with period T, or equivalently with dominant frequency $1/T$, and we expect the superposition

in (25) to also have this dominant frequency. So we can see that Fourier analysis of the image coordinate trajectory contains information about the frequency spectrum of the trajectory in the world.

7.1.2 Projective Camera

The situation is more complicated when the full projective model is used, which takes into account the effects of perspective. If we assume, without loss of generality, that the world coordinate system is centered at the camera and aligned with the camera's coordinate system, and that the focal length is one, then the projection of the trajectory into the image is formed according to:

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \frac{1}{Z(t)} \begin{pmatrix} X(t) \\ Y(t) \end{pmatrix}. \tag{26}$$

The first derivative of the signal $x(t)$ is:

$$\frac{d}{dt}\{x(t)\} = \dot{x}(t) = \frac{1}{Z(t)}\dot{X}(t) - \frac{X(t)}{Z^2(t)}\dot{Z}(t). \tag{27}$$

The Fourier transform of the the $\dot{x}(t)$ signal is then:

$$\begin{aligned} \mathscr{F}(\dot{x}(t)) =& \mathscr{F}\left(\frac{1}{Z(t)}\right) \otimes \mathscr{F}\left(\dot{X}(t)\right) \\ & - \mathscr{F}\left(\frac{X(t)}{Z^2(t)}\right) \otimes \mathscr{F}\left(\dot{Z}(t)\right), \end{aligned} \tag{28}$$

where $\otimes$ represents convolution. As before, we expect $\dot{X}(t)$ and $\dot{Z}(t)$ to have dominant frequencies at $1/T$. The signals $\frac{1}{Z(t)}$ and $\frac{X(t)}{Z^2(t)}$ we expect to have significant frequency components at $1/T$, but to also have dominant peaks at very low frequencies due to the translational components of the motion (note that these two signals are constructed from the position of the point, not its velocity). So in many cases these convolutions are expected to preserve the dominant spectral component $1/T$ in the Fourier transform of $\dot{x}(t)$ (because of the large low-frequency components of the position signals), and to also make copies of this dominant peak at harmonics of the frequency $1/T$. Specifically, we expect this to be true when the very low-frequency components of the position signals contain more of the signal's power than the periodic component.

Note that the expressions above are quite similar for the case of the $y(t)$ signal in image coordinates. In fact, in practice the dominant frequency can be extracted by analyzing the Fourier transform of linear combinations $cos(\phi)\dot{x}(t) + sin(\phi)\dot{y}(t)$. The angle ϕ represents the angle of the line in the image onto which the image coordinate trajectory is projected. Some projections may better preserve the dominant frequency $1/T$ than others, and this frequency can be extracted by observing the spectra of multiple of these projections.

7.2 Sparsity-Weighted Period Estimation

In the previous subsection it was demonstrated that some or all of the temporal frequency information (depending on which camera model is used) about the 3D velocity is preserved when it is projected into image coordinates, even though the pure periodicity may be distorted by the projection. This seems to indicate that it makes sense to use the Fourier transform to extract the dominant frequency. In this subsection, we outline the actual algorithm used to extract the dominant frequency, which has been shown to perform well in practice.

Once a point undergoing periodic motion has been tracked, its image coordinate trajectory can be thought of as a curve through x-y-t-space, where, x and y are image coordinates, and t is time. If one were to plot the gradients of the $x(t)$ and $y(t)$ signals separately (i.e., the image-coordinate velocities, $\dot{x}(t)$ and $\dot{y}(t)$), it might be observed that one of these signals appears to be "more periodic" than the other. This can be understood in terms of Equation (28), since one of the projections ($x(t)$ or $y(t)$) may have preserved the periodicity of the original trajectory better than the other.

In fact, $\dot{x}(t)$ and $\dot{y}(t)$ are just two examples of projections of the x-y-t velocity onto planes parallel to the t-axis. In general, the image-coordinate velocity signal can be projected onto any such plane (given by a rotation ϕ about the t-axis), yielding the projected velocity signal:

$$v_\phi(t) = \begin{pmatrix} \cos(\phi) \\ \sin(\phi) \end{pmatrix}^T \begin{pmatrix} \dot{x}(t) \\ \dot{y}(t) \end{pmatrix}. \tag{29}$$

In most cases, we find that some of these projections preserve much of the original (world coordinate) signal's periodicity, while others are more severely affected by the projection into image coordinates. This can be understood as a generalization of Equation (28), where here more projections of the image velocity signal are taken into account. To asses this, we consider the Power Spectral Density (PSD) of the Fourier transform of each projection, denoted as

$P_\phi(f)$, which is a function of frequency. Specifically, we are interested in the *sparsity* of the PSD signals $P_\phi(f)$. Velocity projections $v_\phi(t)$ that preserve more of the original periodicity have typically been observed to be more sparse, consisting of one dominant frequency component which corresponds to the period of the motion in the world, with other frequency coefficients remaining relatively small.

To estimate the period of motion we formulate an optimization problem, where the goal is to identify the dominant frequency over the aggregate of all the PSD signals $P_\phi(f)$. The aggregate PSD is computed as:

$$P_A(f) = \sum_\phi w_\phi P_\phi(f), \tag{30}$$

where the weights w_ϕ are proportional to the sparsity of $P_\phi(f)$. One common measure of sparsity is the inverse of the ℓ_p-norm, with $0 < p \leq 1$ (Donoho, 2006; Hurley and Rickard, 2008). The weights in our case are given by:

$$w_\phi = \frac{1}{\|P_\phi(f)\|_p}, \tag{31}$$

where we treat the discrete PSD $P_\phi(f)$ as a vector. In practice, we have found the $\ell_{0.5}$-norm to work well; however, the performance is not greatly tied to the specific value of p. Note that, before the weights (31) are computed, each PSD signal $P_\phi(f)$ is normalized by its maximum values. Finally, the fundamental frequency of motion (i.e., the inverse of the period of motion) is given by the optimizer of the problem:

$$\arg\max_f \sum_\phi w_\phi P_\phi(f). \tag{32}$$

8 Experimental Results

In this section we present experimental results which demonstrate the proposed reconstruction techniques and explore their feasibility. §8.1 uses synthetic data, while the remainder of the experiments use real periodic motions, mostly from human movement.

8.1 Noise Sensitivity Analysis

Two different formulations for reconstructing periodic trajectories in 3D have been presented. The first is based on minimizing the reprojection error (§4), and requires an iterative local optimization procedure to obtain the minimizer. The second formulation (§5) is based on minimizing a 3D geometric error. Recall that this cost function can be minimized through a single SVD.

In this experiment we aim to assess how sensitive each of these cost functions is to noise. We test this in a Monte-Carlo fashion, using synthetically generated periodic trajectories with increasing levels of noise. Specifically, the data consisted of four different periodic trajectories, shown in Figure 3. Synthetic noise was added to these trajectories in image coordinates by drawing random samples from an isotropic two-dimensional Gaussian distribution, with variance ranging from 10^{-3} to 6.3 pixels-squared. At each level of noise covariance, Monte Carlo simulations were repeated 5 times for each of the 4 trajectories.

The results of the simulation are shown in Figure 4. Reconstruction was performed using the reprojection error method, as well as the 3D geometric error. The 3D geometric error cost function was used with second-order regularization, with three different weighting coefficients. In each case we measured the reconstruction error between the actual trajectory and its reconstruction by finding the alignment that minimizes the Orthogonal Procrustes Distance (Schonemann, 1966; Eggert et al., 1997). Figure 4 shows the mean of the reconstruction error over all simulations at each noise level. Note that the units of the reconstruction error hold little meaning, since we know that reconstruction is accurate only up to a scale. Examples of reconstructions of the circular spiral trajectory are shown in Figure 5 for four different levels of noise, plotted with the actual trajectory. Note that each of the three dimensions is plotted against time, in order to better visualize the differences. As can be seen, the quality of the reconstruction deteriorates as the amount of noise increases.

Several important observations can be made from Figure 4. First, we see that the reprojection error method (blue line) and the 3D geometric error method with no regularization (green line) show similar performance. This is quite an important observation, since the 3D geometric error cost function is significantly easier to minimize. This result indicates that these two cost functions can be used interchangeably.

For very low levels of noise, both the reprojection error and unregularized geometric error methods achieve reconstruction errors lower than the regularized cost functions. This is expected, since regularization terms combat noise, but may also distort the reconstruction slightly. In fact, with zero noise

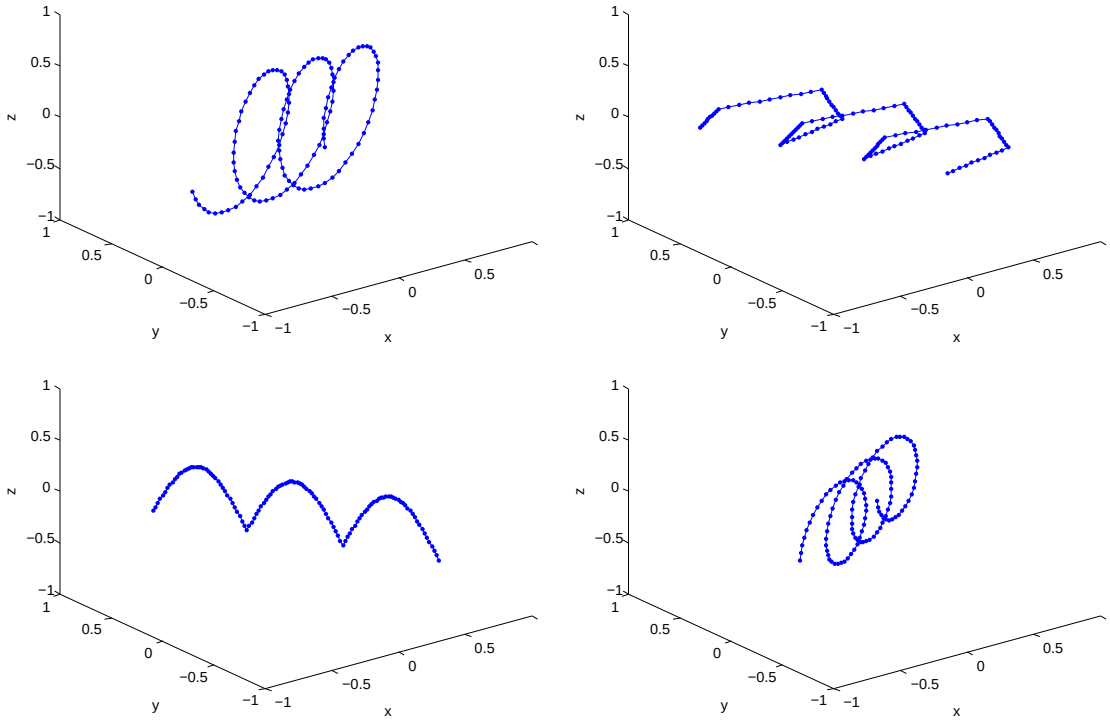

Figure 3: The four synthetic trajectories used in the sensitivity analysis, including (row-wise from top-left) circular spiral, rectangular spiral, arches, and a point on a wheel.

added, both the reprojection error and unregularized geometric error methods achieve reconstruction errors that are near zero, to machine precision. Another important observation is that, when noise variance reaches approximately 10^{-1}, the geometric error with a regularization coefficient of 0.05 (red line) begins to perform better than the other cost functions. This is because the regularization combats the effects of the noise, helping it achieve more smooth reconstructions. Finally, we note that the geometric error method with a larger regularization coefficient (black line) generally performs worse than the other techniques. This shows that with too much regularization, the distortion in the reconstruction may outweigh the benefits of the smoothing properties.

8.2 Accuracy of Period Estimation

§7 described an algorithm for estimating the period of a trajectory from only its appearance in image coordinates. Here we examine the accuracy of the proposed technique. The data used in this experiment consisted of 68 trajectories from 2 different people performing different periodic motions, including walking, running, sideways shuffling and marching. Points of interest were tracked in image coordinates using an automatic colored marker-based tracker, so that significant noise was introduced. The subjects were instructed only to move as evenly as possible between two points.

For each trajectory, the period of motion was estimated automatically using the technique described in §7 based on sparsity-weighted PSDs. For comparison, the periods were also labeled manually. Since the period was not always constant over an entire sequence, the upper and lower bounds of the stride-time were manually labeled for each trajectory. A plot of the automatic estimates, along with the manually labeled upper and lower bounds, is shown in Figure 6. The sequences were reordered to produce a smoother plot. As can be seen, the automatic estimate typically falls in or very near the manual bounds.

In addition, for each sequence we also measured how far the period estimate was outside of the manually labeled bounds. If the estimate was between the upper and lower bounds, a distance of zero was counted. This distance can be thought of as a period estimation error. A histogram of the period estima-

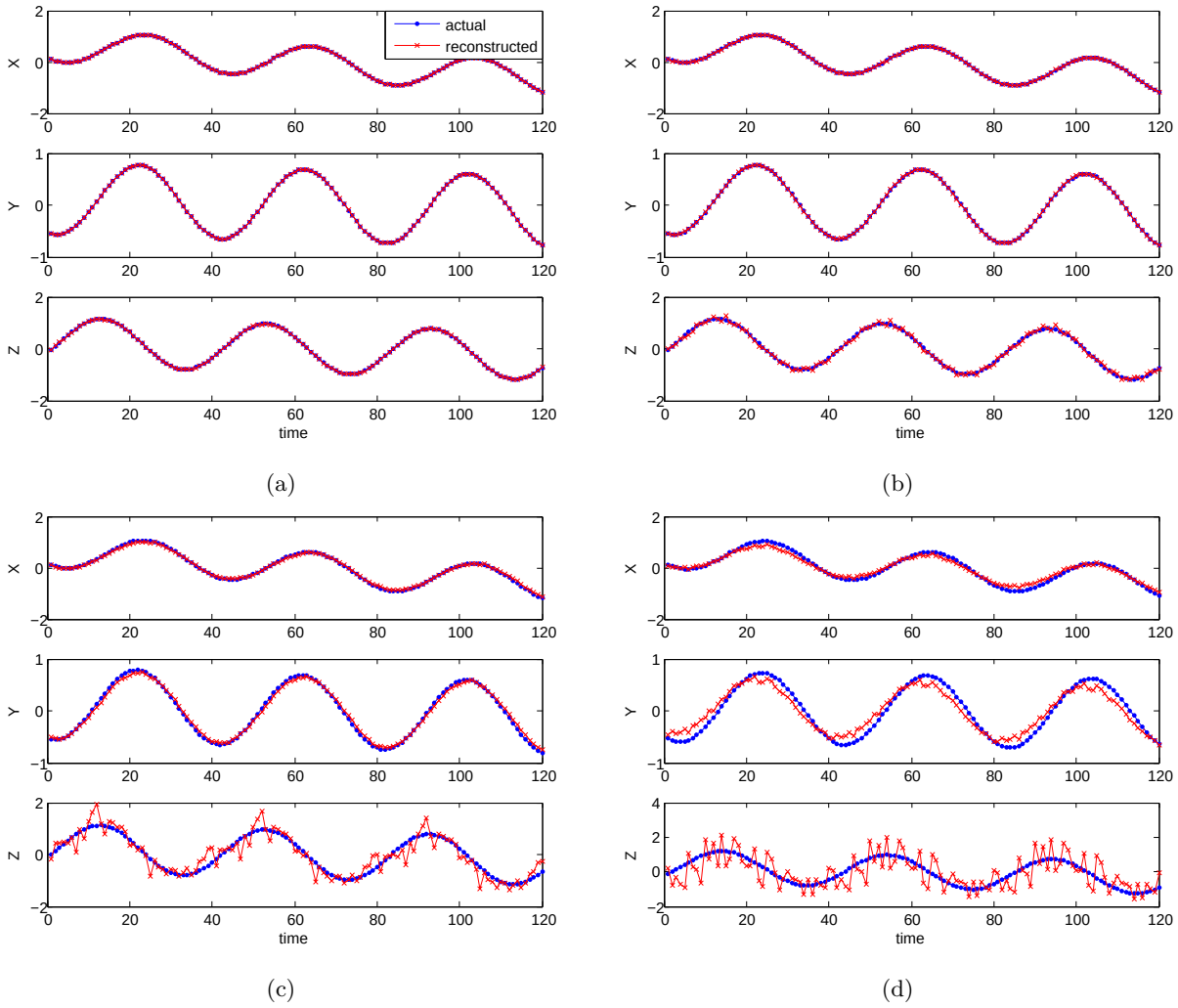

Figure 5: Four reconstructions of the circular spiral trajectory, with noise covariance (a) 10^{-3}, (b) 0.028, (c) 0.92, and (d) 5.3 pixels squared.

tion errors can be seen in Figure 7, which shows that the period was estimated to within 0.1 seconds for a vast majority of the sequences (98.5%). Note that the actual periods for these sequences were between 0.6 and 1.1 seconds, as can be seen from Figure 6.

8.3 Reconstruction Accuracy

Here we present several examples of real periodic motion, and in each case examine the accuracy of the reconstruction.

8.3.1 Vehicle Wheel

In this experiment we consider the motion of a point on the wheel of a vehicle as it drives with constant velocity. A snapshot from the video can be seen in Figure 8, with the apparent trajectory of the point in image coordinates superimposed. For this trajectory it was possible to collect ground truth data regarding its motion in 3D, since it is clear that a point on the wheel moves in a vertical plane in world coordinates, and that this plane intersects the ground plane at known positions. Extrinsic camera calibration was used to ascertain the ground truth.

The 3D reconstructed trajectory of one period of motion, along with the entire ground truth trajec-

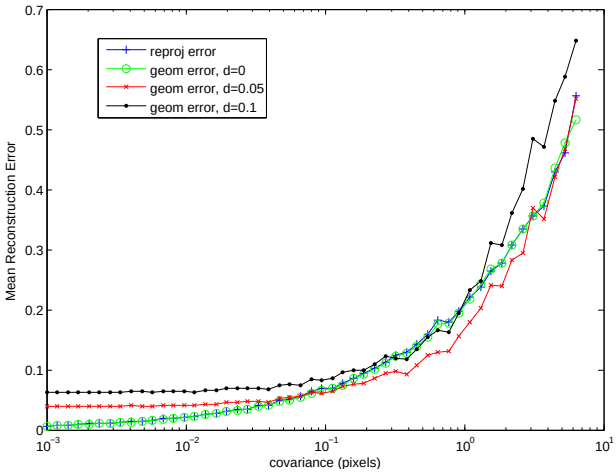

Figure 4: Results of noise sensitivity simulations. Mean reconstruction error vs. noise variance. For the 3D geometric error, the value of the second-order regularization coefficient is denoted by "d".

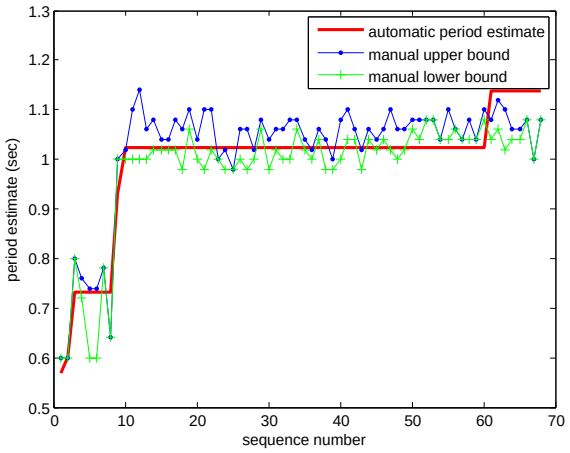

Figure 6: Automatic period estimates for all 68 sequences, along with the manually labeled upper and lower bounds.

tory, can be seen in Figure 9. Note that the reconstructed trajectory closely matches the actual positions of the samples. These results are summarized quantitatively in Table 1, where we have used the extrinsic calibration to represent these results on a global coordinate system, with the XY-plane corresponding to the ground, and Z corresponding to world height, with Y representing the depth from the camera on the ground plane. The errors are relatively small when compared with the distance of the object

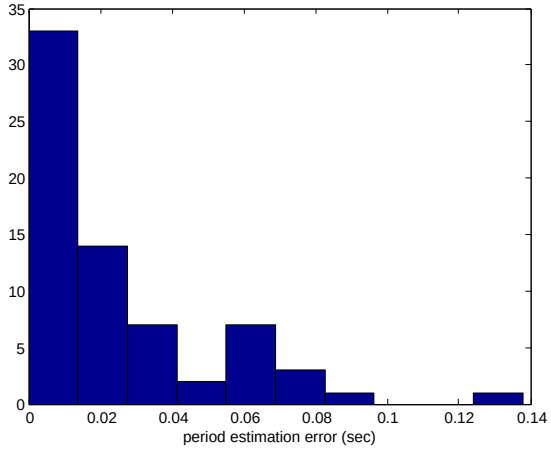

Figure 7: Histogram of period estimation errors.

from the camera (on the order of 600cm). Errors are higher in the Y-coordinate (depth), since it is more sensitive to noise and inaccuracies in the camera calibration.

8.3.2 Hand and Foot of Walking Person

Next we consider the periodic motions of points on a person's hand and foot as he walks. An image from the video is shown in Figure 10. As can be expected from real human motion, these trajectories contain significant noise and small deviations from true periodicity. Accurate ground truth data was collected here using a motion capture system.

Figure 11 shows the 3D reconstructed trajectories

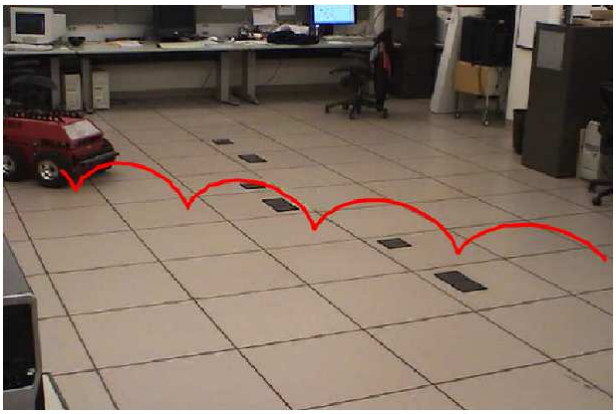

Figure 8: Trajectory of a point on the wheel of a vehicle as it drives.

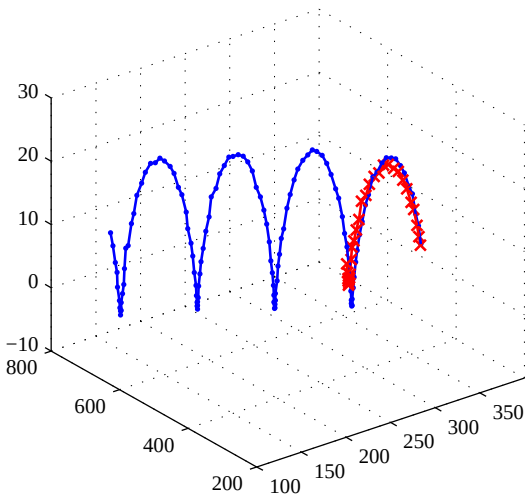

Figure 9: The 3D reconstruction of the trajectory of a point on a wheel (red crosses), along with the approximate ground truth (blue circles). Axes are shown in centimeters.

Table 1: Absolute errors of the 3D reconstruction of the wheel trajectory for one period of motion. The object's distance from the camera was on the order of $600cm$.

	Mean	Std. Dev.
X (cm)	4.21	2.09
Y (cm)	5.65	3.15
Z (cm)	2.62	1.51
Euclidean (cm)	7.69	3.72

of one period for both the hand and foot, along with the full ground truth trajectories. As before, the reconstructed periods closely match the ground truth. The reconstruction errors for both trajectories are given in Table 2, where the errors are again in cm. These errors are small, given that the distance of the person from the camera was on the order of $300cm$.

8.3.3 Hand of Walking Person

Another example of periodic motion is from a point on a person's hand as he walks. An image from the video used here is shown in Figure 12. As can be seen from the apparent trajectory superimposed, this motion contains significant noise and small deviations from true periodicity.

Figure 12 shows the reprojection into image coordi-

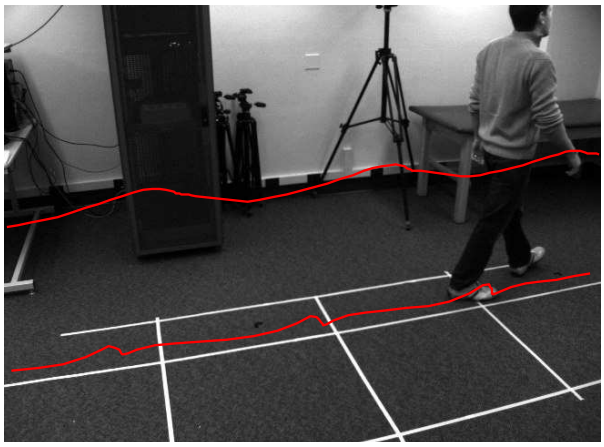

Figure 10: The trajectories of a person's hand and foot as he walks.

Table 2: Absolute errors of the 3D reconstruction for the hand and foot trajectories over one period of motion. The person's distance from the camera was on the order of $300cm$.

	Foot		*Hand*	
	Mean	Std. Dev.	Mean	Std. Dev.
X (cm)	0.79	0.99	4.76	2.51
Y (cm)	3.16	4.67	2.49	1.58
Z (cm)	2.26	2.98	0.98	1.10
Euclidean (cm)	4.02	5.59	5.59	2.92

nates of the 3D trajectory estimated by our method for one period, along with the actual trajectory in the image for several periods. In order to show that the estimated trajectory is also accurate in 3D world coordinates, we use it to estimate the stride of the person as exhibited by the displacement of his hand from one period to the next. This was compared to his actual stride, which was approximated by observing the points at which his foot intersects the ground plane in the video, again making use of an extrinsic camera calibration. The results are summarized in Table 3. Displacement in the Z direction (corresponding to height) was zero.

8.3.4 Foot of Running Person on Stairs

For the next experiment we consider a point on the foot of a person running down a set of stairs. Notice that in this case the person's foot displaces also in the vertical direction from one period to another. A snapshot from the video is shown in Figure 13. The point tracked in this case was on the tip of the person's left foot. Figure 13 shows the estimated trajectory repro-

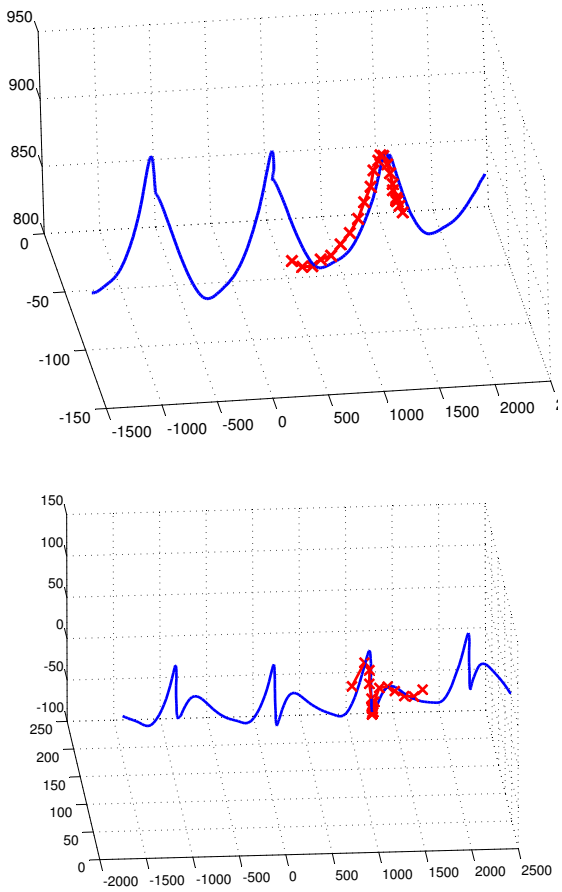

Figure 11: The 3D reconstruction of the hand (top) and foot (bottom) trajectories (red crosses), along with the ground truth (blue circles). Axes are shown in millimeters.

jected into image coordinates for one period, along with the actual trajectory in image coordinates.

In order to demonstrate the accuracy of the reconstruction in 3D, we use our estimated trajectory to infer the dimensions of the stairs on which the person runs, and compare these to the actual dimensions of the staircase. Specifically, we compared the stair height with $|\Delta_{Z_T}|/2$, and the stair depth with $|\Delta_{Y_T}|/2$, and found the dimensions inferred from our estimate to be quite accurate. As before, we have used extrinsic calibration to represent the reconstruction on a world-centered coordinate system, with the axes aligned in this case with the stairs. These results are summarized in Table 4, which shows that the stair dimensions were estimated relatively accurately.

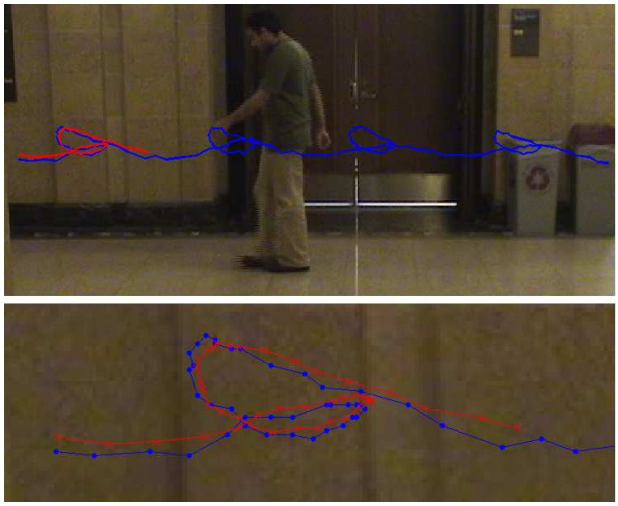

Figure 12: The trajectory of a person's hand as he walks.

Table 3: Comparison of the stride estimated from the person's hand with the actual stride of his foot. The distance of the person from the camera was on the order of $800cm$.

	Estimated	Actual
Δ_{X_T} (cm)	97.44	87.60
Δ_{Y_T} (cm)	9.09	1.82
Stride Length (cm)	97.87	87.61

8.4 View-Invariance

The purpose of this experiment is to show that different reconstructions of a given motion are very similar to one another, independent of the angle from which the motion is viewed. To this end, we analyzed several videos of humans walking in a straight line. The videos contained 34 walking sequences from two different people, filmed simultaneously from two very different viewing angles. Each trajectory was reconstructed separately from both views.

An example of a single walking sequence, recon-

Table 4: Dimensions of the stairs inferred from the estimate of the trajectory, compared with the actual dimensions of the stairs. The distance of the person from the camera was on the order of $700cm$.

	Estimated	Actual
Stair Height (cm)	14.56	15.875
Stair Depth (cm)	25.52	33.02

Figure 13: The trajectory of a person's foot as he runs down stairs.

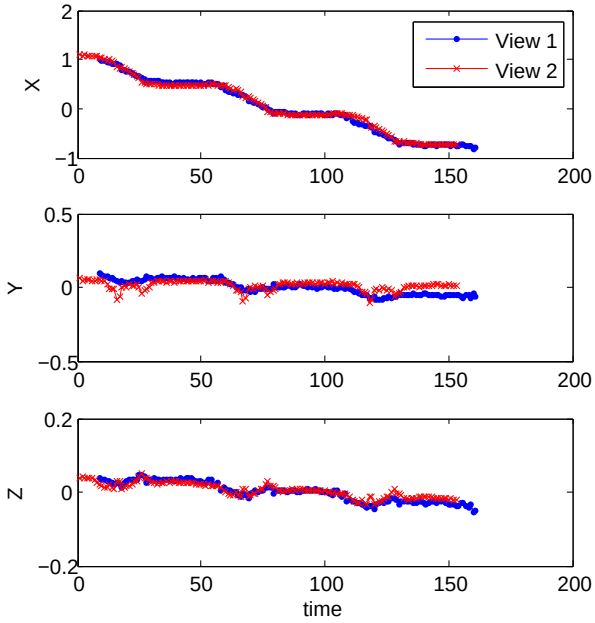

Figure 14: Reconstructions of the same motion from two different views. Images from the original videos are shown in Fig. 1.

structed independently from two views, is shown in Figure 14. Images from the original videos are shown in Figure 1. Notice that even though the trajectory of the tracked point is the same in both cases, its appearance in image coordinates changes dramatically with the viewing angle. Figure 14 shows plots of the reconstructions from the two views superimposed, where the reconstructions have been aligned via translation, rotation and scaling. Note that they are plotted on the new axes, as described above, where the first axis is the principal component. Three periods of the reconstructions are shown, plotted against time on the horizontal axis. As can be seen, the two reconstructions match each other very well, even over multiple periods.

Similar analysis was performed for each of the 34 motion sequences, and the distance between each corresponding pair of reconstructions was computed – the so-called inter-view reconstruction error. Since the reconstructions are on an arbitrary scale, the distance between them is computed as a fraction of the total displacement over the three periods. We found that the average inter-view reconstruction error was 8.7% of the total displacement, which shows that the corresponding reconstructions from different views match each other quite well.

9 Conclusions and Future Work

In this paper we have explored the possibility of reconstructing periodic trajectories in 3D using only their appearances in image coordinates from stationary monocular views. This was motivated by the importance of periodicity as a cue in several computer vision tasks on the one hand, and the lack of view-invariance exhibited by most existing techniques on the other hand. The general paradigm envisioned is one in which a motion is first reconstructed based on its image-coordinate appearance, and then any subsequent analysis is performed in 3D, independent of the original viewing angle.

Two different formulations for obtaining such reconstructions have been developed, based on minimization of the reprojection error and a 3D geometric error, respectively. We have also presented theoretical results, derived from an analogy with multi-view geometry, which guarantee solvability under certain, easily achievable geometric conditions. In addition, we have introduced a novel technique for estimating the period of motion based on its apparent trajectory in image coordinates. Through careful experimentation we have analyzed the robustness to noise of these proposed formulations, and demonstrated their accuracy and view-invariance.

In future research, it would be interesting to further explore the applicability of these reconstruction tech-

niques in tasks involving human motion. Along these lines, it may become necessary to develop additional algorithmic tools to handle the specific complexities of natural human motion. This includes reconstruction techniques for motions that deviate significantly from pure periodicity, and tools for handling multiple points on an articulated body.

Acknowledgements

This material is based upon work supported in part by the Department of Homeland Security, the Center for Transportation Studies and the ITS Institute at the University of Minnesota, the Minnesota Department of Transportation, the U.S. Army Research Laboratory and the U.S. Army Research Office under contract #911NF-08-1-0463 (Proposal 55111-CI), and the National Science Foundation through grants #IIS-0219863, #CNS-0224363, #CNS-0324864, #CNS-0420836, #IIP-0443945, #IIP-0726109, and #CNS-0708344.

References

M. Allmen and C.R. Dyer. Cyclic motion detection using spatiotemporal surfaces and curves. In *Proc. Int'l. Conf. Pattern Recognition*, volume 1, pages 365–370, 1990. 3

S. Belongie and J. Wills. Structure from periodic motion. In *Proc. Int'l. Workshop on Spatial Coherence for Visual Motion Analysis*, pages 16–24, May 2004. 3

S.P. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004. 6

A. Briassouli and N. Ahuja. Extraction and analysis of multiple periodic motions in video sequences. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 29 (7):1244–1261, July 2007. 3

C.J. Cohen, L. Conway, and D. Koditschek. Dynamical system representation, generation, and recognition of basic oscillatory motion gestures. In *Proc. Int'l. Conf. Automatic Face and Gesture Recognition*, pages 60–65, 1996. 3

R. Cutler and L.S. Davis. Robust real-time periodic motion detection, analysis, and applications. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(8):781–796, August 2000. 3

J. Davis, A. Bobick, and W. Richards. Categorical representation and recognition of oscillatory motion patterns. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, volume 1, pages 628–635, 2000. 3

D.L. Donoho. Compressed sensing. *IEEE Trans. Information Theory*, 52:1289–1306, 2006. 11

D.W. Eggert, A. Lorusso, and R.B. Fisher. Estimating 3-d rigid body transformations: A comparison of four major algorithms. *Machine Vision and Applications*, 9:272–290, 1997. 11

A. Elgammal and C.-S. Lee. Inferring 3D body pose from silhouettes using activity manifold learning. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, volume 2, pages 681–688, 2004. 1

A. Fossati, E. Arnaud, R. Horaud, and P. Fua. Tracking articulated bodies using generalized expectation maximization. In *Proc. CVPR Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment*, 2008. 2

H. Fujiyoshi, A.J. Lipton, and T. Kanade. Real-time human motion analysis by image skeletonization. *IEICE Trans. Information and Systems*, E87-D(1):113–120, 2004. 3

R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2 edition, 2003. 6, 7, 8, 9

N. Hurley and S. Rickard. Comparing measures of sparsity. In *Proc. IEEE Workshop on Machine Learning for Signal Processing*, pages 55–60, 2008. 11

J. Little and J. Boyd. Describing motion for recognition. In *Proc. Int'l. Symposium on Computer Vision*, pages 235–240, 1995. 3

F. Liu and R.W. Picard. Finding periodicity in space and time. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 1998. 3

J. Nocedal and S.J. Wright. *Numerical Optimization*. Springer, 1999. 5

D. Ormoneit, M.J. Black, T. Hastie, and H. Kjellstrom. Representing cyclic human motion using functional analysis. *Image and Vision Computing*, 23:1264–1276, 2005. 3

X. Orriols and X. Binefa. Classifying periodic motions in video sequences. In *Proc. IEEE Int'l. Conf. Image Processing*, volume 1, September 2003. 3

R. Polana and R. Nelson. Detection and recognition of periodic, nonrigid motion. *Int'l. J. Computer Vision*, 23(3):261–282, 1997. 3

D. Ramanan, D.A. Forsyth, and A. Zisserman. Tracking people by learning their appearance. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 29:65–81, 2007. 2

E. Ribnick and N. Papanikolopoulos. Estimating 3D trajectories of periodic motions from stationary monocular views. In *Proc. European Conf. Computer Vision*, 2008. 2, 5

E. Ribnick and N. Papanikolopoulos. View-invariant analysis of periodic motion. In *Proc. IEEE/RSJ Int'l. Conf. Intelligent Robots and Systems*, 2009. 2

E. Ribnick, S. Atev, and N. Papanikolopoulos. Estimating 3D positions and velocities of projectiles from monocular views. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 31:938–944, 2009. 1

S. Schaible and J. Shi. Fractional programming: the sum-of-ratios case. *Optimization Methods and Software*, 18: 219–229, 2003. 5

P. Schonemann. A generalized solution of the orthogonal procrustes problem. *Psychometrika*, 31:1–10, 1966. 11

S.M. Seitz and C.R. Dyer. View-invariant analysis of cyclic motion. *Int'l. J. Computer Vision*, 25:1–23, 1997. 3

L. Sigal and M. J. Black. Measure locally, reason globally: Occlusion-sensitive articulated pose estimation. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2006. 2

P.-S. Tsai, M. Shah, K. Keiter, and T. Kasparis. Cyclic motion detection for motion based recognition. *Pattern Reocognition*, 27(12):1591–1603, 1994. 3

Z. Zhang and N.F. Troje. 3D periodic human motion reconstruction from 2D motion sequences. *Neural Computation*, 19:1400–1421, 2007. 2