Estimating 3D Trajectories of Periodic Motions from Stationary Monocular Views

Evan Ribnick and Nikolaos Papanikolopoulos

University of Minnesota, USA
{ribnick,npapas}@cs.umn.edu

Abstract. We present a method for estimating the 3D trajectory of an object undergoing periodic motion in world coordinates by observing its apparent trajectory in a video taken from a single stationary camera. Periodicity in 3D is used here as a physical constraint, from which accurate solutions can be obtained. A detailed analysis is performed, from which we gain significant insight regarding the nature of the problem and the information that is required to arrive at a unique solution. Subsequently, a robust, numerical approach is proposed, and it is demonstrated that the cost function exhibits strong local convexity which is amenable to local optimization methods. Experimental results indicate the effectiveness of the proposed method for reconstructing periodic trajectories in 3D.

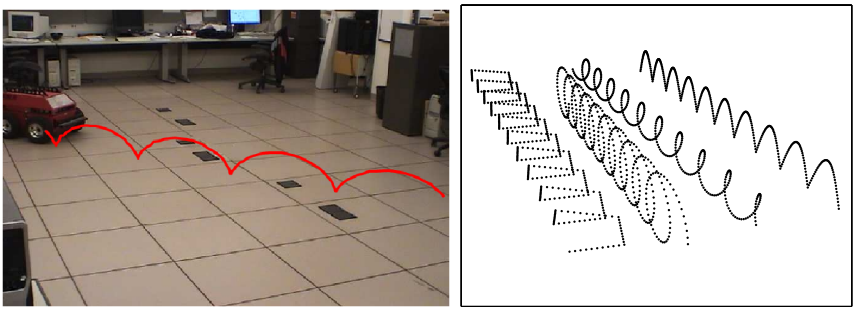
1 Introduction

Periodic motion occurs frequently from both natural and mechanical sources. For example, consider the motion of a person or animal running, or a point on the wheel of a car as it drives. The ability to analyze periodic motions when they appear in video is important in many applications. However, the problem of inferring 3D information about a periodic motion is still an open problem that has not been addressed in the literature.

In this paper we propose a method for reconstructing the 3D trajectory of an object undergoing periodic motion in the world given its apparent trajectory in a video from a single stationary view. For example, Fig. 1(a) shows the trajectory of the point on the wheel of a vehicle as it moves at a constant speed. While clearly this trajectory is periodic in 3D world coordinates, its appearance in the image is distorted by the perspective projection, and as such is not periodic in image coordinates. However, we understand intuitively that information regarding the 3D trajectory of the object in world coordinates is embedded in its apparent trajectory in the image. The goal here is to infer this 3D information in order to estimate the trajectory of the object in world coordinates.

The ability to reconstruct a periodic motion in 3D given only its appearance in a single stationary view has potential applications in many domains. Activity recognition, gait analysis, surveillance applications, and motion analysis for athletic training or physical therapy might all benefit from techniques such as the one proposed here. For example, in gait analysis/recognition, it may be possible to better characterize a gait by examining it in 3D rather than in image coordinates, since a description of the motion in world coordinates will be independent of both the intrinsic and extrinsic parameters of the camera, and will hence be more portable.

D. Forsyth, P. Torr, and A. Zisserman (Eds.): ECCV 2008, Part III, LNCS 5304, pp. 546–559, 2008.



(a) (b)

Fig. 1. Examples of periodic trajectories

Periodicity is used here as a physical constraint on the motion in the world, which is exploited in the formulation. We require throughout that the period of the motion is known, which can be estimated automatically using one of several existing methods (e.g., [4]). In our analysis we find that the problem is under-constrained in general, and that an additional constraint is needed in order to arrive at a unique solution. This additional constraint, along with the initial constraint of periodicity in the world, make it possible to estimate the 3D position of each sample in the trajectory, and hence to reconstruct it as completely as possible in 3D at the given sampling rate. The perspective projection is fully accounted for in the formulation; in fact, the problem addressed here could not be solved in an orthographic system.

We also require a calibrated camera, where both the intrinsic and extrinsic parameters are known. This was a design choice made by the authors, since without a calibrated camera it would not be possible to obtain an exact Euclidean reconstruction. With an uncalibrated camera, rather, it would only be possible to obtain at best an affine reconstruction, which would severely limit the usefulness of the proposed method.

For this method we assume perfect periodicity, which might not be valid in many real-world applications. However, this assumption is often acceptable over relatively short windows of time for a given motion. Furthermore, experimental results indicate the effectiveness of this approach for real motions, even under this assumption. Development of methods that account for fluctuations in period will be the focus of future work.

This work is unique in that it is the first (to the best of the authors' knowledge) to reconstruct a periodic trajectory in 3D. This is done using video from a single stationary camera, from which it is more difficult to infer 3D information than from multi-camera systems. However, since it is often difficult to obtain video of a motion from multiple views in practice, examining such a problem is worthwhile.

Existing work related to periodic motion in video has focused mainly on detection and analysis in image coordinates. For example, in [4] and [8] periodic motion is detected via sequence alignment, where the similarity between snapshots of a foreground object is computed and tested for periodicity. In both methods the effects of imaging are taken into account by considering the change in the apparent size of the object in the image as it moves relative to the position of the camera. Another approach is to perform frequency domain analysis of image intensities in order to detect periodic motion in the image ([9]); this approach is used in [2] to detect multiple periodic motions

simultaneously. Similarly, the authors of [11] develop a method for recognition of different classes of periodic motion based on a frequency domain analysis of image intensities.

In other work, the assumption of strict periodicity has been relaxed in order to consider the case of cyclic ([13]) or oscillatory ([3]) motion in the image. In [5], a method is proposed for learning cylindrical manifolds for cyclic motion, where the radial parameter accounts for the inherent repetitiveness.

Finally, in [1], a method is presented which can be used to infer the 3D structure of an object undergoing periodic motion by considering multiple snapshots of the object separated by one period in time as if they were simultaneous views of the same object. These multiple views are used to perform geometric inference using techniques from multiview geometry.

2 Problem Formulation

The specific problem that we wish to solve can be stated more formally as follows. Given image coordinate samples from the trajectory of an object that is undergoing periodic motion in the world, our goal is to estimate the 3D world position of each sample in the most recent period of motion. An image coordinate sample is defined here as a single point in image coordinates corresponding to the apparent position of the object at the time of sampling. Note that when the 3D position of every sample during one period is known, the periodic motion is fully characterized (at the given sampling rate), since each period is simply a repetition of the preceding period. Before formalizing the problem, we first introduce some notation and define the quantities we wish to estimate.

Periodic motion is taken to mean any motion in 3D world coordinates that is periodic in velocity $\mathbf{v} \triangleq (\dot{X}, \dot{Y}, \dot{Z})$:

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

where T is the period of the motion and n is an integer. Notice that this definition differs from that in some previous work (e.g.,, [4]), 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 a person's foot as he/she walks. In fact, the translational component is necessary in order to solve the problem formulated here, since the technique is based on 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 Sec. 3.1.

For motion that is periodic in velocity we can write:

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

where $\mathbf{\Delta}_{\mathbf{p}_T} \triangleq (\Delta_{X_T}, \Delta_{Y_T}, \Delta_{Y_T})$ is the displacement per period of the point, which is constant over any period of length T. This can be used to describe the 3D displacement between any two samples that are separated by exactly one period in time.

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.

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. These are reverse time indices, so that t_k^i is taken to mean the time of the k-th most recent sample in period i, which is the i-th most recent period. Given these definitions and equations (2) and (3), each sample in world coordinates can be written in terms of the 3D position at the most recent sample, $\mathbf{p}_0^0 = (X_0^0, Y_0^0, Z_0^0)$, as follows:

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

Here $\delta_{\mathbf{p}_k}$ is the 3D displacement between the samples t_k^i and t_0^i, which is constant for all periods i. Expressing each sample in terms of these quantities will prove to be useful, since these are the quantities that need to be estimated in order to fully characterize one period of motion in 3D.

The method described in this paper relies on an accurate camera calibration, which can be obtained using techniques similar to those described in [6]. As a result of the calibration, we obtain the projection matrix $\mathbf{P}_{\mathrm{c}} = \mathbf{A}_{\mathrm{c}}\mathbf{I}_{3\times4}\mathbf{T}_{\mathrm{c}} \in \mathbb{R}^{3\times4}$, where $\mathbf{A}_{\mathrm{c}} \in \mathbb{R}^{3\times3}$ describes the camera's intrinsic parameters, $\mathbf{I}_{3\times4} \in \mathbb{R}^{3\times4}$ is a non-square identity matrix, and $\mathbf{T}_{\mathrm{c}} \in \mathbb{R}^{4\times4}$ describes the camera's extrinsic parameters. $\mathbf{P}_{\mathrm{c}}$ is used to project points from world coordinates into the image, using homogeneous coordinates, according to:

$$\begin{pmatrix} u & v & h \end{pmatrix}^T = \mathbf{P}_{\mathrm{c}} \begin{pmatrix} X & Y & Z & 1 \end{pmatrix}^T, \tag{5}$$

where the image coordinates can be obtained by dividing by the homogeneous coordinate: $x = u/h$, $y = v/h$. So, given the location of the object in the world at time t_k^i, denoted in homogeneous coordinates as $\mathbf{q}_k^i = \begin{pmatrix} X_k^i & Y_k^i & Z_k^i & 1 \end{pmatrix}^T$, the image coordinates of this point can be written as follows:

$$x_k^i = \frac{\mathbf{P}_{\mathrm{c}(1)}\mathbf{q}_k^i}{\mathbf{P}_{\mathrm{c}(3)}\mathbf{q}_k^i}, \quad y_k^i = \frac{\mathbf{P}_{\mathrm{c}(2)}\mathbf{q}_k^i}{\mathbf{P}_{\mathrm{c}(3)}\mathbf{q}_k^i}, \tag{6}$$

where $\mathbf{P}_{\mathrm{c}(i)}$ is the i-th row of $\mathbf{P}_{\mathrm{c}}$. Using (4) we can write $\mathbf{q}_k^i$ in terms of the quantities for which we wish to solve:

$$\mathbf{q}_k^i = \begin{pmatrix} X_0^0 - i\Delta_{X_T} - \delta_{X_k} \\ Y_0^0 - i\Delta_{Y_T} - \delta_{Y_k} \\ Z_0^0 - i\Delta_{Z_T} - \delta_{Z_k} \\ 1 \end{pmatrix}. \tag{7}$$

3 Existence of Solution

Here we consider the ideal, noise-free version of the problem in order to analyze when it is possible to solve it uniquely. As such, we proceed by splitting the problem into

two subproblems, and analyzing each one separately. First, we consider the problem of estimating the 3D position of the most recent sample (X_0^0, Y_0^0, Z_0^0) and the per period displacement $(\Delta_{X_T}, \Delta_{Y_T}, \Delta_{Z_T})$ by taking only samples at times separated by exactly one period from t_0^0 — that is, samples at times t_0^i. We find that this problem is underconstrained in general, and therefore an additional constraint must be added in order to arrive at a unique solution. Then, given (X_0^0, Y_0^0, Z_0^0) and $(\Delta_{X_T}, \Delta_{Y_T}, \Delta_{Z_T})$, we focus on the problem of estimating the 3D positions of the other samples in the most recent period by solving for the inter-sample displacements $(\delta_{X_k}, \delta_{Y_k}, \delta_{Z_k})$. In each case the subproblem is formulated as a linear system in order to make some its inherent properties more evident. Subsequently, in Section 4, these problems will be reformulated in a more robust way in order to account for measurement noise.

3.1 Position at t_0^0 and Per Period Displacement

Only samples at times t_0^i are to be considered for this subproblem. There are six quantities that we wish to estimate ((X_0^0, Y_0^0, Z_0^0) and $(\Delta_{X_T}, \Delta_{Y_T}, \Delta_{Z_T})$), so clearly at least three image coordinate samples are needed, since each image sample will provide two constraints (6) on the system. These three samples will are denoted $(x_0^{i_0}, y_0^{i_0})$, $(x_0^{i_1}, y_0^{i_1})$, and $(x_0^{i_2}, y_0^{i_2})$, respectively. Using these three samples, we can get three pairs of equations (i.e., six equations) of the form (6). Note that here $(\delta_{X_k}, \delta_{Y_k}, \delta_{Z_k}) = 0$, since the samples we consider are taken at integer multiples of the period T from time t_0^0. By rearranging the terms, these six equations can be written as a linear system:

$$\mathbf{A}_6^x \mathbf{X}_6 = \mathbf{B}_6^x, \tag{8}$$

where $\mathbf{X}_6 \triangleq \begin{pmatrix} X_0^0 & Y_0^0 & Z_0^0 & \Delta_{X_T} & \Delta_{Y_T} & \Delta_{Z_T} \end{pmatrix}^T \in \mathbb{R}^6$ and $\mathbf{A}_6^x \in \mathbb{R}^{6\times 6}$ is the coefficient matrix. It is important to note that in order to write the equations as a linear system, we have multiplied the projection equations (6) through by their denominator. Originally, the system of equations was such that each of the samples had equal weight. This multiplication has, in effect, weighted each image sample by the denominator term. For the ideal, noise-free case on which we are currently focusing, this re-weighting will have no effect, and is convenient for the sake of analysis. However, when measurement noise is present, a more robust approach will be necessary, and will be formulated later.

Upon analysis it becomes clear that system (8), with six equations and six unknowns, is in fact underdetermined when the image samples are collinear. This will always be the case for periodic motion when the samples are separated by exactly one period in time, since they will be collinear in the world. The result is that $\mathrm{rank}(\mathbf{A}_6^x) = 5$ in general.

As such, an additional constraint is needed. For all the experiments performed here, we add this constraint by requiring the value of Z_0^0 (the Z-coordinate of the object's position at time t_0^0) to be known. Note that this condition is not expected to be overly restrictive in practice, since Z_0^0 can often be inferred automatically based on *a priori* knowledge about the motion in the world. For example, if it is known that the object undergoing periodic motion is the foot of a walking person, then it may be inferred that $Z = 0$ at the lowest point on the trajectory. Additionally, the problem could be formulated in a similar way if either the X- or Y-coordinate was known instead.

With this additional constraint there are now five remaining quantities to be estimated: (X_0^0, Y_0^0) and $(\Delta_{X_T}, \Delta_{Y_T}, \Delta_{Z_T})$. Since Z_0^0 is known, the third column of $\mathbf{A}_6^x$ can be eliminated (these coefficients will be absorbed by the vector on the other side of the equation), and one of the equations in the system is discarded. The remaining system of five unknowns and five equations is written as follows:

$$\mathbf{A}_5^x \mathbf{X}_5 = \mathbf{B}_5^x, \tag{9}$$

where $\mathbf{X}_5 \triangleq \begin{pmatrix} X_0^0 & Y_0^0 & \Delta_{X_T} & \Delta_{Y_T} & \Delta_{Z_T} \end{pmatrix}^T \in \mathbb{R}^5$ and $\mathbf{A}_5^x \in \mathbb{R}^{5\times 5}$.

Recall now that the original coefficient matrix $\mathbf{A}_6^x$ contained the image coordinates of the samples from which the equations were written: $(x_0^{i_0}, y_0^{i_0})$, $(x_0^{i_1}, y_0^{i_1})$, and $(x_0^{i_0}, y_0^{i_0})$. If we now replace these image coordinate values with their expressions in terms of 3D quantities using the projection (6), we find that the rowspace of $\mathbf{A}_5^x$ is spanned by the rows of the following matrix:

$$\begin{pmatrix} \Delta_{Z_T} & 0 & 0 & 0 & -\frac{D_{123}X_0^0 - D_{234}}{D_{123}} \\ 0 & \Delta_{Z_T} & 0 & 0 & -\frac{D_{123}Y_0^0 - D_{134}}{D_{123}} \\ 0 & 0 & 0 & 0 & -\frac{D_{123}Z_0^0 - D_{124}}{D_{123}} \\ 0 & 0 & \Delta_{Z_T} & 0 & -\Delta_{X_T} \\ 0 & 0 & 0 & \Delta_{Z_T} & -\Delta_{Y_T} \end{pmatrix}, \tag{10}$$

where D_{abc} is the determinant of the 3×3 matrix formed by the a-th, b-th, and c-th columns of the projection matrix $\mathbf{P}_c$. Note that $D_{123} \neq 0$, since it is the determinant of $\mathbf{A}_c\mathbf{R}$, the product of the camera's intrinsic parameter and rotation matrices, both of which are nonsingular by definition.

It will be possible to obtain a unique solution to (9) when $\mathrm{rank}(\mathbf{A}_5^x) = 5$. Clearly $\mathbf{A}_5^x$ will be full rank only when all the rows of the matrix (10) are nonzero. This, in effect, imposes physical conditions under which the problem will become unsolvable. The first important observation is that in order for all the rows of (10) to be nonzero, some per period displacement is required (in other words, some translation is necessary in order for the problem to be solved). In addition, (10) imposes conditions on X_0^0, Y_0^0, and Z_0^0. However, if any of these conditions are satisfied and $\mathbf{A}_5^x$ becomes singular, the problem can simply be re-parameterized around a new point (X_0^0, Y_0^0, Z_0^0) such that these conditions are no longer satisfied.

3.2 Position at t_k^0

For this subproblem we consider samples at times t_k^i. The goal now is to estimate the 3D displacement between samples t_0^0 and t_k^0, $(\delta_{X_k}, \delta_{Y_k}, \delta_{Z_k})$, given the quantities that are known from solving the previous subproblem (i.e., (X_0^0, Y_0^0, Z_0^0) and $(\Delta_{X_T}, \Delta_{Y_T}, \Delta_{Z_T})$ are known). Note that samples at times t_k^i can also be expressed in terms of these previously estimated quantities. Since there are three quantities to be estimated, clearly at least three equations are needed, and therefore we consider two image coordinate samples, which are taken at times t_k^i and t_k^j, respectively. Since these two samples result in four equations (recall each sample results in two equations as in (6)), we discard one of them to obtain the three equations. As before, we multiply all the equations through by their denominators and write them as a linear system:

$$\mathbf{A}_3^\delta \delta_k = \mathbf{B}_3^\delta, \tag{11}$$

where $\delta_k \triangleq \begin{pmatrix}\delta_{X_k} & \delta_{Y_k} & \delta_{Z_k}\end{pmatrix}^T \in \mathbb{R}^3$ and $\mathbf{A}_3^\delta \in \mathbb{R}^{3\times 3}$ is the coefficient matrix. As mentioned earlier, writing these nonlinear equations as a linear system will introduce nonuniform weights on the samples, but is convenient for analysis in the noise-free case.

In order to obtain a unique solution for the system (11), matrix $\mathbf{A}_3^\delta$ must be nonsingular. Recall that one of the four equations resulting from the two image samples was discarded in order to form this 3×3 system. If one of the x-samples was discarded, we find that:

$$\det(\mathbf{A}_3^\delta) = -(y_k^j - y_k^i) D_{123}. \tag{12}$$

Similarly, if one of the y-samples was discarded:

$$\det(\mathbf{A}_3^\delta) = -(x_k^j - x_k^i) D_{123}. \tag{13}$$

It is guaranteed that at least one of the two determinants above will be nonzero, since the samples (separated by one period) must be distinct and D_{123} is known to be nonzero. As such, this subproblem can always be solved as long as the per period displacement $\mathbf{\Delta}_{\mathbf{P}T}$ is nonzero. This analysis is identical for each sample in the period, t_k^i, and therefore the trajectory of the object in 3D can be completely characterized.

4 Trajectory Estimation

In the previous section it was shown that when one additional physical constraint is added to the problem, a unique solution for the 3D position of each sample in the most recent period can be obtained in most cases by solving systems of linear equations derived from samples of the apparent trajectory in the image. However, in practice, the solution from just two or three image samples is sensitive to measurement noise and is often unreliable. Furthermore, dividing the problem into subproblems and solving each one separately ignores some of the interdependencies between variables, and as such is not as robust to noise.

As such, a more robust approach to estimating the 3D position of each sample in the most recent period of motion is to use all of the available samples of the trajectory, rather than just the minimum number required for nonsingularity. In addition, here we estimate all parameters simultaneously instead of solving each subproblem separately. We propose a numerical approach for performing the estimation, in which a cost function is minimized.

4.1 Cost Function

As mentioned above, here we estimate all variables of interest simultaneously (recall that Z_0^0 is known). If we have image samples from M periods with N samples per period, there are a total of $5+3(N-1)$ variables. These include the unknown coordinates of the most recent sample (X_0^0, Y_0^0), the per period displacement $(\Delta_{X_T}, \Delta_{Y_T}, \Delta_{Z_T})$,

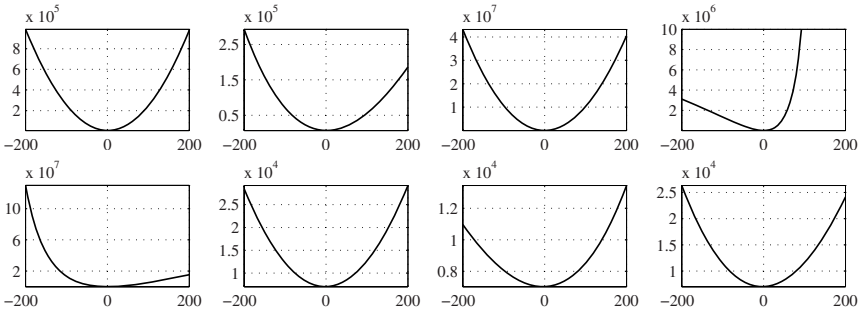


Fig. 2. Examples of cross sections of the cost function $F(\hat{\mathbf{X}})$ with significant measurement noise. The first eight dimensions are shown. Cross sections are plotted against the offset from the optimal value, which are shown in centimeters.

and the displacements within each period $(\delta_{X_k}, \delta_{Y_k}, \delta_{Z_k})$, $k \in [1, N-1]$. We concatenate all of these quantities into a single vector of variables, and denote the current estimate as $\hat{\mathbf{X}} \in \mathbb{R}^{(5+3(N-1))}$. The image samples are denoted as (x_k^i, y_k^i), $k \in [1, N-1]$, $i \in [1, M-1]$. Then, for all $M \times N$ samples, we wish to minimize the squared reprojection error:

$$\hat{\mathbf{X}}^* = \arg\min F(\hat{\mathbf{X}}), \tag{14}$$

where:

$$F(\hat{\mathbf{X}}) \triangleq \sum_{i=0}^{M-1} \sum_{k=0}^{N-1} \left\| \begin{pmatrix} x_k^i \\ y_k^i \end{pmatrix} - \begin{pmatrix} \hat{x}_k^i \\ \hat{y}_k^i \end{pmatrix} \right\|_2^2, \tag{15}$$

and $\| \cdot \|_2$ is the ℓ_2 norm. The reprojections in image coordinates $(\hat{x}_k^i, \hat{y}_k^i)$ are functions of the current estimate of the variables, $\hat{\mathbf{X}}$, calculated using the projection equations (6).

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

$$F(\hat{\mathbf{X}}) = \sum_{i=0}^{M-1} \sum_{k=0}^{N-1} \frac{\hat{\mathbf{X}}^T \mathbf{A}_k^i \hat{\mathbf{X}} + (\mathbf{B}_k^i)^T \hat{\mathbf{X}} + C_u^i}{\hat{\mathbf{X}}^T \mathbf{D}_k^i \hat{\mathbf{X}} + (\mathbf{E}_k^i)^T \hat{\mathbf{X}} + F_k^i}, \tag{16}$$

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, $F(\hat{\mathbf{X}})$ 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 [12]. Furthermore, the individual subfunctions in the summation are nonconvex themselves, since they are ratios of quadratic functions. However, we have found that, in practice, $F(\hat{\mathbf{X}})$ is locally convex around its optimal solution, and that this convex region is typically large, even in the presence of significant measurement noise. An example of some of the cross sections of $F(\hat{\mathbf{X}})$ can be seen in Fig. 2. Recall that $F(\hat{\mathbf{X}})$ is a function on $\mathbb{R}^{(5+3(N-1))}$; cross sections on only the first eight dimensions are shown here. In the next section it will be demonstrated that local optimization methods can be used to reliably converge to the optimal solution in the presence of measurement noise.

One important point is that the cost function (15) properly accounts for noise, since it gives equal weight to all samples. This makes it more robust to noise than the linear

systems in Sec. 3, which were formulated by multiplying the projection equations (6) by their denominator, hence giving nonuniform weights to the image samples. Recall that the linear systems were useful only for analyzing the problem in the ideal, noise-free case in order to determine when it is possible to arrive at a unique solution.

Although the present formulation is numerical instead of analytic, the analysis in Sec. 3 still gives us significant insight as to the nature of the problem in general. The numerical estimation procedure proposed here provides more information from which to solve the problem, which should only improve our ability to arrive at an appropriate solution. This will be shown subsequently. Furthermore, it is important to note that the optimization here aims to obtain an optimal estimate directly. This is in contrast to techniques such as those proposed in [7], which optimize convex relaxations of their cost functions in order to approximate the optimal solution for similar geometric problems.

4.2 Initial Solution

The linear systems formulated in Sec. 3 are not robust to noise, since they introduce nonuniform weights to the image samples and divide the problem into parts. However, we use them to obtain an initial solution to the cost function (15). Specifically, we wish to use all of the available samples in order to solve the first subproblem (9), and the second subproblem (11) for each sample in the period. As such, the coefficient matrix $\mathbf{A}_5^x$ and vector $\mathbf{B}_5^x$ for the first subproblem, and $\mathbf{A}_3^\delta$ and $\mathbf{B}_3^\delta$, $k \in [1, N-1]$, are augmented with additional rows corresponding to the additional samples that are to be used for estimation. This results in the matrices being non-square. The initial solution to the optimization is then obtained by solving the linear least-squares cost function $\|\mathbf{AX} - \mathbf{B}\|_2^2$ in each case. In all experiments performed here, we find that the initial solution found in this manner is within the convex region of the original cost function (15), and therefore local optimization methods can be used to converge to the optimal solution.

5 Experiments

In all of the experiments described here, a variant of Levenberg-Marquardt (LM) [10] was used to perform the minimization of the cost function described in Sec. 4. LM is convenient since it automatically interpolates between quasi-Newton's method in more convex regions, and gradient descent in less convex regions. Convergence was always achieved in a reasonable number of iterations, and in most cases the LM algorithm tended more towards quasi-Newton's method than gradient descent (this is controlled by the LM parameter λ), which supports the assertion that the cost function $F(\hat{\mathbf{X}})$ is locally convex in a large area surrounding the optimal solution. The initial solution was found by minimizing the simple linear least-squares cost functions for each separate subproblem, as described previously.

5.1 Results on Synthetic Data

As a proof of concept, some initial experiments were performed on synthetic data in order to judge the effectiveness of the proposed method. For these experiments, four

different synthetic trajectories were generated, sampled, and projected into image coordinates using a user-specified projection matrix. Evaluation of the performance of the proposed method is particularly convenient in this case, since the 3D world position of each sample is known exactly for synthetic trajectories. The projections of the synthetic trajectories in the image are shown in Fig. 1(b), where the trajectories shown in the figure have been rotated/translated in world coordinates in order for them to fit into a single figure. As can be seen, there are various types of trajectories represented, with several different displacement patterns.

In all of these experiments the periodic trajectories were accurately reconstructed in 3D world coordinates. The errors (Euclidean distances) between the actual positions and the estimates from our method were on the order of $10^{-12}cm$ in all cases. In other words, since these errors were measured in centimeters, the overall error was negligible. Furthermore, the value of the cost function always converged close to zero (on the order of 10^{-25}, which is less than machine precision). Convergence occurred in a reasonable number of iterations. These experiments clearly show that the proposed method is feasible, and that the theoretical results discussed earlier do apply in practice.

5.2 Results on Real Data

Next we present results that illustrate the performance of the proposed technique on real data. The trajectories on which the algorithm was tested each represent a different type of periodic motion. The data used for testing can be obtained by contacting the authors. For each experiment the point of interest was tracked manually in image coordinates, and supplied as input to our algorithm.

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 Fig. 1(a), 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.

The 3D reconstructed trajectory of one period of motion, along with the entire ground truth trajectory, can be seen in Fig. 3. Note that the reconstructed trajectory closely matches the actual positions of the samples. These results are summarized quantitatively in Table 1. The errors are relatively small when compared with the distance of the object 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.

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 Fig. 4(a). 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[1].

Fig. 5 shows the 3D reconstructed trajectories of one period for both the hand and foot, along with the full ground truth trajectories. As before, the reconstructed periods

[1] Thanks to Prof. David Nuckley for his assistance in collecting motion capture data.

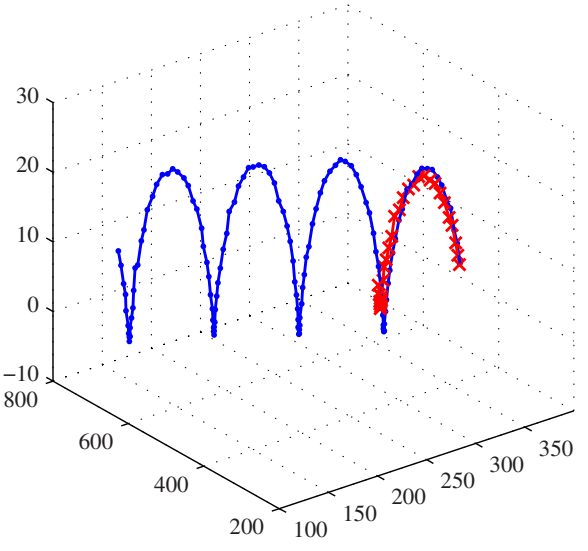


Fig. 3. The 3D reconstruction of the trajectory of a point on a wheel (red crosses), along with the approximate ground truth (blue circles). Axes are shown in centimeters. This figure is best viewed in color.

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

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 $300cm$.

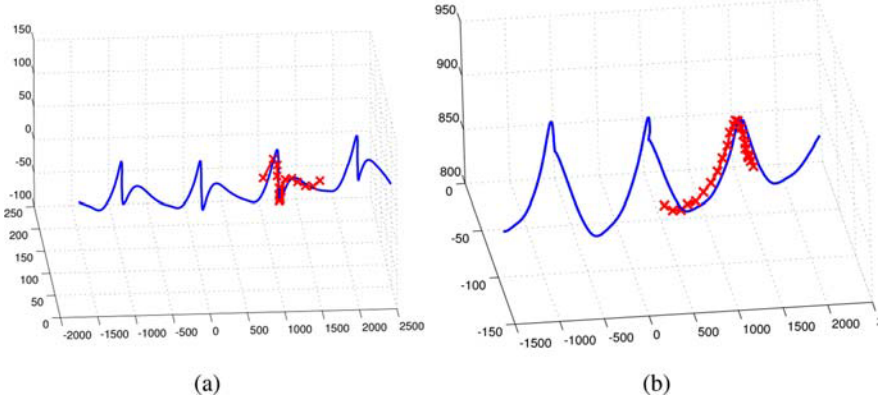
	Foot Mean	*Foot* Std. Dev.	*Hand* Mean	*Hand* Std. Dev.
X (cm)	0.79	0.99	4.76	2.51
Y (cm)	3.16	4.67	2.49	1.58
Z (cm)	2.26	2.98	0.98	1.10
Euclidean (cm)	4.02	5.59	5.59	2.92

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 $300cm$.

Foot of Running Person on Stairs. For the final 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

(a) (b)

Fig. 4. (a) Trajectories of a person's hand and foot as he walks. (b) Trajectory of a person's foot as he runs down stairs. Both the actual trajectory in image coordinates (blue) and the reprojection of the trajectory for one period estimated by our method (red) are shown. This figure is best viewed in color.



(a) (b)

Fig. 5. The 3D reconstruction of the foot (a) and hand (b) trajectories (red crosses), along with the ground truth (blue circles). Axes are shown in millimeters. This figure is best viewed in color.

Table 3. 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

displaces also in the vertical direction from one period to another. A snapshot from the video is shown in Fig. 4(b). Since the person in the video steps on the ground at the bottom of the stairs, it could simply be inferred that $Z_0^0 = 0$. The point tracked in this case was on the tip of the person's left foot. Fig. 4(b) shows the estimated trajectory reprojected 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. These results are summarized in Table 3.

6 Conclusions and Future Work

We have presented a method that can be used to accurately reconstruct a periodic trajectory in 3D based on samples of the trajectory in image coordinates. Analysis of the problem showed that it is under-constrained in general, and the additional constraint that the value of Z_0^0 is known was added. It was demonstrated that a numerical approach can be used to perform the estimation, and that the cost function has been found to be locally convex around the optimal solution. Experimental results illustrated that this technique can be used to accurately infer 3D information, even in the presence of noise and imperfect periodicity.

In light of the fact that real motions are often not perfectly periodic, in future work we plan to develop methods for performing 3D inference in the more general case of cyclic motion. In this case it may be necessary to estimate the instantaneous frequency of the motion at each point.

Acknowledgements

This work has been supported 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, and the National Science Foundation through grants #IIS-0219863, #CNS-0224363, #CNS-0324864, #CNS-0420836, #IIP-0443945, #IIP-0726109, and #CNS-0708344.

References

1. Belongie, S., Wills, J.: Structure from Periodic Motion. In: Proc. Int'l. Worksh. Spatial Coherence for Visual Motion Anal., pp. 16–24 (May 2004)
2. Briassouli, A., Ahuja, N.: Extraction and Analysis of Multiple Periodic Motions in Video Sequences. IEEE Trans. Pattern Anal. Mach. Intel. 29(7), 1244–1261 (2007)
3. Cohen, C.J., et al.: Dynamical System Representation, Generation, and Recognition of Basic Oscillatory Motion Gestures. In: Proc. Int'l. Conf. Automatic Face and Gesture Recognition, pp. 60–65 (October 1996)
4. Cutler, R., Davis, L.S.: Robust Real-Time Periodic Motion Detection, Analysis, and Applications. IEEE Trans. Pattern Anal. Mach. Intel. 22(8), 781–796 (2000)
5. Dixon, M., et al.: Finding Minimal Parameterizations of Cylindrical Image Manifolds. In: Proc. Comp. Vis. and Pattern Recog. Worksh., p. 192 (June 2006)
6. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press, Cambridge (2003)
7. Kahl, F., Henrion, D.: Globally Optimal Estimates for Geometric Reconstruction Problems. Int'l. J. Comp. Vis. 74(1), 3–15 (2007)

8. Laptev, I., et al.: Periodic Motion Detection and Segmentation via Approximate Sequence Alignment. In: Proc. IEEE Int'l. Conf. Comp. Vis., vol. 1, pp. 816–823 (October 2005)
9. Liu, F., Picard, R.: Finding Periodicity in Space and Time. In: Proc. IEEE Int'l. Conf. Comp. Vis., pp. 376–383 (January 1998)
10. Nocedal, J., Wright, S.J.: Numerical Optimization. Springer, New York (1999)
11. Polana, R., Nelson, R.C.: Detection and Recognition of Periodic, Nonrigid Motion. Int'l. J. Comp. Vis. 23(3), 261–282 (1997)
12. Schaible, S., Shi, J.: Fractional Programming: the Sum–of–Ratios Case. Optimization Methods and Software 18, 219–229 (2003)
13. Seitz, S.M., Dyer, C.R.: View-Invariant Analysis of Cyclic Motion. Int'l. J. Comp. Vis. 25(3), 231–251 (1997)