Shortened version appearing in *Proc. CVPR*, Seattle, WA, 1994.

Affine Invariant Detection of Periodic Motion

Steven M. Seitz
seitz@cs.wisc.edu

Charles R. Dyer
dyer@cs.wisc.edu

Department of Computer Sciences
University of Wisconsin
Madison, WI 53706

Technical Report #1225
June 1994

Abstract

Current approaches for detecting periodic motion assume a stationary camera and place limits on an object's motion. These approaches rely on the assumption that a periodic motion projects to a set of periodic image curves, an assumption that is invalid in general. Using affine-invariance, we derive necessary and sufficient conditions for an image sequence to be the projection of a periodic motion. No restrictions are placed on either the motion of the camera or the object. Our algorithm is shown to be provably-correct for noise-free data and is extended to be robust with respect to occlusions and noise. The extended algorithm is evaluated with real and synthetic image sequences.

The support of the National Science Foundation under Grant No. IRI-9220782 is gratefully acknowledged.

1 Introduction

Many real-life motions can be characterized as having a periodic component. For instance, most human locomotory motions (e.g., walking, running, skipping, shuffling) can be decomposed into a repetitive motion (e.g., running in place) and a net translatory component. To determine periodicity information, the periodic component needs to be isolated. The problem is compounded with a moving camera because changes in viewpoint may alter the projection of a periodic motion.

In this paper we describe a technique for determining periodicity information from image sequences that is invariant with respect to (1) changes in the position, orientation, and scale of the moving object, and (2) changes in viewpoint. The key observation is that effects due to both (1) and (2) can be modeled by affine transformations so we can make use of recent results in affine-invariance [3, 8].

Periodicity is remarkable in that it can be detected without taking into account the structure of the objects producing the periodic motion. Therefore periodicity information can be detected at an early stage and used by higher level processes of a motion-analysis system. A problem in current motion recognition systems [2, 5] is temporal alignment; it is difficult to match two motions that are out of phase or have different periods. The latter difficulty is resolved by detecting the period at an early stage in the visual analysis process.

Periodic motions have a regular structure that can be exploited to determine some important qualities of the object, including its 3D shape, identity, and motion. In fact, many techniques designed for the analysis of rigid-body motion can be naturally extended to cope with non-rigid periodic motions. The key observation is that each configuration of a periodically moving object repeats in the sense that it reappears in each cycle, perhaps in a different position or orientation. When these corresponding configurations are grouped together, they can be analyzed using affine or rigid-body techniques (e.g., shape-from-motion, recognition) since they are all rigid transformations of the same configuration. For periodic motions, the problem of determining corresponding images reduces to determining the period.

In addition to being interesting in its own right, the period of a motion is often linked to important properties that may otherwise be difficult to determine. Examples include the relation of heart-rate to activity and fitness levels, the period of a hand on a clock to units of time, and the period of locomotion to velocity. To elucidate the last example, imagine that somebody is running towards you from a distance and you wish to determine their velocity. If the person is far away, it is difficult to judge their velocity based on cues like looming. On the other hand, if the runner's period can be measured, the approximate velocity can be determined based on the correlation of stride frequency to net translational velocity. In fact, if the aforementioned correlation is known, the period provides a measure of velocity that is *invariant* with respect to the relative positions and orientations of the observer and the runner. In particular, the period can be used to determine velocity without recovering depth.

The remainder of the paper is structured as follows. Section 2 discusses related work. Section 3 formally defines the concept of periodic motion. Section 4 describes affine-invariance and matching, and lays the theoretical groundwork for the rest of the paper. Section 5 presents an algorithm for detecting a periodic motion and determining its period, and Section 6 presents some experimental results.

2 Related Work

Several researchers have investigated ways of measuring periodicity from image-sequences [1, 4, 9]. Allmen and Dyer [1] described an approach for detecting periodicity under orthographic projection for an object that does not rotate in depth. They used the curvature scale space of point trajectories to detect repeating patterns of curvature maxima and hence infer a period. Polana and Nelson [4] presented a method for detecting periodic motions using Fourier transforms of several point trajectories. In theory, the period of the motion could be detected as well by averaging the fundamental frequencies of the point trajectories, although the authors indicated that determining the period in this way was unreliable. Tsai and Shah [9] described a similar technique, using Fourier transforms of curvature values, where the period was determined from a single point trajectory.

Each of the previous approaches relied on the assumption that periodic point trajectories of an object appear periodic in projection. Unfortunately, this assumption is invalid when the camera is allowed to move (even if the camera is restricted to move parallel to the image plane). Equivalently, the assumption is invalid when an object is allowed unrestricted rigid movement during the course of a periodic motion. For example, consider the chaotic motion of a flying bat. The bat exhibits a motion that is periodic relative to a bat-centered coordinate frame. However, no single point on the bat will appear to move in a periodic manner in a camera-centered frame. Detecting the period of the flapping wings of the bat requires considering the *relative* motion of different points on the object.

3 Periodic Motions

We find it convenient to represent a motion $\mathbf{M}(t)$ of n point trajectories as a time-varying 3 by n matrix. The ith column of $\mathbf{M}(t)$ represents the 3D position of the ith point at time t.

We call a motion $\mathbf{M}$ *purely-periodic* if it repeats with period p, i.e.,

$$\mathbf{M}(t+p) = \mathbf{M}(t) \tag{1}$$

holds for some constant $p > 0$ and all times t in a given time domain. We call the smallest such constant p the *period*.

Examples of purely-periodic motions include a rotating wheel, a spinning top, waving gestures, and a person running in place. Notice that a purely-periodic motion in $\Re^3$ will produce a projection that is purely-periodic (in $\Re^2$) when the camera is stationary, so no knowledge of the projection process is needed to determine periodicity information. The more challenging cases involve motions that are only partially periodic or involve a moving camera. For example, suppose we allow a runner to move along an arbitrary path, around a track for instance. Intuitively, the runner still exhibits a motion that repeats, in some sense, although the motion may no longer satisfy Eq. (1). Notice, however, that the runner's motion *is* purely-periodic in a reference frame that moves with the runner. The apparent motion of the runner with respect to a stationary camera can be decomposed into two component motions, that of running in place, and the rigid-body motion induced by a moving reference frame[1].

[1] Others have referred to this decomposition in terms of "relative" and "common" motion, respectively.

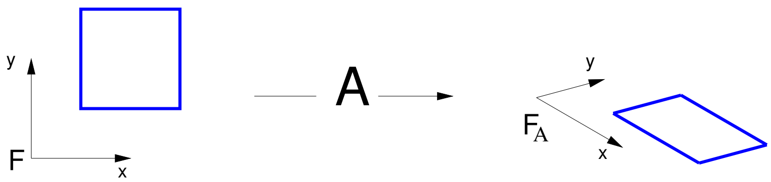

Figure 1: An affine transformation, **A**, of the coordinate axes transforms a square to a parallelogram. A parallelogram can be interpreted as a square in *some* reference frame.

Eq. (1) can be generalized to accommodate a moving reference frame by introducing a class of allowable transformations that does not affect the periodicity. Specifically, note that any change in reference frame can be modeled by a 3D affine transformation of the scene (i.e., a rotation and translation) with respect to the camera frame. The converse does not hold unless the notion of reference frame is generalized to include all affine transformations (e.g., scaling, bending, and shearing motions). The use of affine reference frames [3, 10] identifies affine transformations with reference frames so that any affine transformation of the scene can be explained by a change in reference frame (see Fig. 1). Therefore, a motion can be considered periodic in *some* reference frame (possibly moving with respect to the camera) if the apparent motion of the scene is periodic modulo affine transformations.

We say that a motion $\mathbf{M}$ is *affinely-periodic* if

$$\mathbf{M}(t) = \mathbf{A}(t) \circ \mathbf{C}(t) \tag{2}$$

for some purely-periodic motion $\mathbf{C}$ and some time-varying affine transformation $\mathbf{A}$.

Because the projection process is not invertible, we cannot hope to detect all possible 3D periodic motions: a projected motion may appear periodic even though the 3D motion that generated it was not, and the converse is also possible. The problem is even worse when the motion is sampled in time (in an image sequence). The best we can hope to do is to determine if the projected sampled motion *could have* been produced by a 3D periodic motion. For clarity we will use the term *still* and the notation $\mathbf{F}_i$ to denote a projection (expressed as a 2 by n matrix) of a motion $\mathbf{C}(i)$.

We call a sequence of stills *apparently-periodic* if the stills are the orthographic[2] projection of a time-sampled, affinely-periodic motion. Notice that the above definition doesn't place any restrictions on the configuration of the camera; the camera parameters may vary from one image to the next.

4 Affine Matching

Our approach is based on the observation that corresponding *stills* from different cycles of a projected, affinely-periodic motion must *match* modulo affine transformations. This notion of matching is actually a necessary and sufficient condition for a motion to be apparently-periodic. In this section we define precisely what is meant by *matching* and demonstrate its relationship to periodic motions. Although the concept of matching extends to other classes of transformations, the results in this section exploit the choice of affine transformations.

4.1 Match Criteria

We say that m 3D shapes *match* if they are all affine transformations of a single shape. Matching relates to periodic motions by the following observation:

3D Match Criterion: *A motion $\boldsymbol{C}$ is affinely-periodic with period p if and only if for each time t there exists a shape $\mathbf{S} \in \Re^3 \times \Re^n$ and a sequence of affine transformations, $\{\mathbf{A}_i\} \subset \Re^3 \times \Re^3$, such that $\mathbf{C}(t + kp) = \mathbf{A}_k \circ \mathbf{S}$ for every integer k.*

In other words, corresponding configurations from different cycles of a periodic motion are affine transformations of the same shape. The observation follows directly from the definitions of Section 3.

Now consider the more difficult problem of determining whether a set of 3D shapes match when the only available data is a projected image of each shape. We say that a set of stills *match* when there exists a single 3D

[2] The use of orthography here is arbitrary. All linear projection models are equivalent in this affine context since the projective transformation can be absorbed by the affine function $\mathbf{A}$ in Eq. (2).

shape which could have produced the stills under affine transformations (i.e., changes in reference frame) and projection. The problem of inferring matches under projection is depicted visually in Fig. 2.

Clearly, corresponding stills from different cycles will be projections of affine transformations of the same object. Here we can use the result, due to Tomasi and Kanade [8], that the matrix formed by the registered concatenation of affinely corresponding stills has rank at most 3. Towards this end, define the measurement matrix of stills $\mathbf{F}_i$ and period p as

$$\mathbf{M}_i^p = \begin{vmatrix} \mathbf{F}_i \\ \mathbf{F}_{i+p} \\ \vdots \\ \mathbf{F}_{i+(k-1)p} \end{vmatrix} \tag{3}$$

The measurement matrix can be registered to eliminate effects due to translation by subtracting each row's centroid from each element in the row [8]. Denote the registered version of $\mathbf{M}_i^p$ as $\hat{\mathbf{M}}_i^p$. By the 3D Match Criterion and [8], $\hat{\mathbf{M}}_i^p$ can be expressed as

$$\hat{\mathbf{M}}_i^p = \mathbf{RS} \tag{4}$$

where $\mathbf{R}$ is $2k$ by 3 and $\mathbf{S}$ is 3 by n. Since $\hat{\mathbf{M}}_i^p$ is the product of two rank 3 (or less) matrices, $\hat{\mathbf{M}}_i^p$ is itself of rank at most 3. Therefore, an affinely-periodic motion produces registered measurement matrices of rank at most 3. Conversely, any registered measurement matrix of rank 3 or less can be decomposed as in Eq. (4) using singular value decomposition [8]. Therefore, we have the following:

Projected Match Criterion: *A sequence of stills, $\mathbf{F}_1, \ldots, \mathbf{F}_m$, is apparently-periodic with period p if and only if $\hat{\mathbf{M}}_i^p$ is of rank at most 3 for $i = 1, \ldots, m$.*

We say that a set of stills $\{\mathbf{F}_{i+jp}\}_{j=0}^{k-1}$ *matches* when $\hat{\mathbf{M}}_i^p$ is of rank 3 or less. In other words, a set of stills match when they can be explained by projections of affine transformations of the same object.

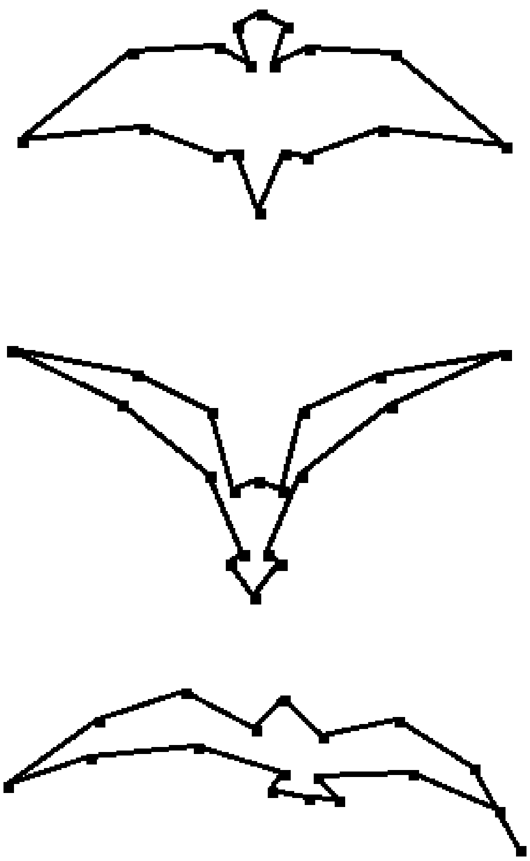

Figure 2: Inferring 3D matches from projected images. The top two images are top and side-views of a stationary bat. The bottom image is a side-view of a bat in a different position. The only bats that match pair-wise are the top two.

4.2 Approximate Matching

Even when a set of stills don't match, we can characterize their relative distance by the amount by which we have to perturb the stills in order to make them match. Define the distance between a set of stills as follows:

$$dist_{\mathcal{A}}(\{\mathbf{F}_{i+jp}\}_{j=0}^{k-1}) = min(\|\mathbf{E}\|_{rms}) \tag{5}$$

where $rank(\hat{\mathbf{M}}_i^p + \mathbf{E}) \leq 3$. $\|\mathbf{E}\|_{rms}$ is the root-mean-squared norm of the matrix $\mathbf{E}$ defined by: $\|\mathbf{E}\|_{rms} = \sqrt{\frac{1}{2kn}\sum_{i,j}\mathbf{E}_{ij}^2}$. In other words, consider all possible ways of additively perturbing a set of stills (by a matrix $\mathbf{E}$) in order to produce a match. Then $dist_{\mathcal{A}}$ is defined to be the norm of the smallest such perturbation.

This definition of distance is not very useful computationally, however, since it implies searching a very large space of possible perturbations. Fortunately, there is a simple way of recharacterizing $dist_{\mathcal{A}}(\{\mathbf{F}_{i+jp}\}_{j=0}^{k-1})$ in terms of the singular values of $\hat{\mathbf{M}}_i^p$:

Theorem 1 (Affine Matching) *$dist_{\mathcal{A}}(\{\mathbf{F}_{i+jp}\}_{j=0}^{k-1}) = \sqrt{\frac{1}{2kn}\sum_{i=4}^{2n}\sigma_i^2}$ where σ_i is the ith singular value of $\hat{\mathbf{M}}_i^p$.*

Proof: Let $\mathbf{X}$ denote $\hat{\mathbf{M}}_i^p$. The singular value decomposition of $\mathbf{X}$ is of the form

$$\mathbf{X} = \mathbf{U}\Sigma\mathbf{V} \tag{6}$$

where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices and the singular values of $\mathbf{X}$ appear along the diagonal of Σ (a $2k$ by n diagonal matrix). The above equation can be re-expressed as

$$\mathbf{X} = \mathbf{U}\tilde{\Sigma}\mathbf{V} + \mathbf{U}\Sigma'\mathbf{V} \tag{7}$$

where $\tilde{\Sigma}$ is of the form

$$\tilde{\Sigma} = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \mathbf{0} \\ & & \sigma_3 & \\ & \mathbf{0} & & \mathbf{0} \end{bmatrix} \tag{8}$$

and Σ' is of the form

$$\Sigma' = \begin{bmatrix} \mathbf{0} & & \mathbf{0} & \\ & \sigma_4 & & \\ \mathbf{0} & & \ddots & \\ & & & \sigma_n \end{bmatrix} \tag{9}$$

Eq. (7) can be expressed as

$$\mathbf{X} = \tilde{\mathbf{X}} + \mathbf{X}' \tag{10}$$

where $\tilde{\mathbf{X}} = \mathbf{U}\tilde{\Sigma}\mathbf{V}$ and $\mathbf{X}' = \mathbf{U}\Sigma'\mathbf{V}$.

$\tilde{\mathbf{X}}$ is the optimal (in an RMS sense) rank-3 approximation to $\mathbf{X}$ [7]. Hence, $\mathbf{X}'$ is the minimal perturbation of $\mathbf{X}$ that produces a match. Therefore, $dist_{\mathcal{A}}(\{\mathbf{X}_i\}_1^m) = \|\mathbf{X}'\|_{rms}$. The latter term is just $\|\Sigma'\|_{rms}$ since $\mathbf{U}$ and $\mathbf{V}$ are orthogonal and the theorem follows. □

The measure $dist_{\mathcal{A}}(\{\mathbf{F}_{i+jp}\}_{j=0}^{k-1})$ gives the average amount (in pixels) necessary to additively perturb the 2D location of each feature in a set of stills in order to produce a perfect match. In the case where there are only two stills, $\mathbf{X}$ and $\mathbf{Y}$, we abbreviate $dist_{\mathcal{A}}(\{\mathbf{X}, \mathbf{Y}\})$ as $dist_{\mathcal{A}}(\mathbf{X}, \mathbf{Y})$.

$dist_{\mathcal{A}}$ has the following properties:

- $dist_{\mathcal{A}}(\{\mathbf{X}_i\}_1^m) = 0$ if and only if the set of stills $\{\mathbf{X}_i\}$ match exactly.
- $dist_{\mathcal{A}}$ is well-behaved with respect to noise since it is defined in terms of feature measurement perturbations. See, for example, [7].
- $dist_{\mathcal{A}}$ is defined in image coordinates and can be directly related to measurement errors.
- $dist_{\mathcal{A}}$ is always zero when less than five features are considered.
- For n features, $dist_{\mathcal{A}}(\mathbf{X}, \mathbf{Y}) = \frac{\sigma_4}{\sqrt{2n}}$ and evaluation cost is $O(n)$ arithmetic operations. For m stills, the evaluation cost is the smaller of $O(nm^2)$ and $O(mn^2)$.

5 An Algorithm

The *Projected Match Criterion* suggests a brute-force algorithm: check every possible value of p, $p = 1, \ldots, \frac{m}{2}$, matching every group of stills having an inter-still spacing of p. If a value of p is found such that $\hat{\mathbf{M}}_i^p$ is of rank 3 or less for every value of i, the motion is apparently-periodic with period p. This scheme requires $O(m^2)$ evaluations of the match function, each requiring $O(nm^2)$ operations for a total cost of $O(nm^4)$.

Although the *Projected Match Criterion* guarantees that the above algorithm is correct for perfect data, the algorithm is not practical in the presence of occlusion and noise. To address the issues of occlusion and loss of features, we match only two stills at a time, requiring only that $\mathbf{F}_i$ and $\mathbf{F}_{i+p}$ have at least five features in common (as opposed to requiring that some number of features be visible throughout the entire sequence). A consequence of matching fewer stills at a time is that absolute correctness is compromised (i.e., it is theoretically possible for m stills to match pair-wise but not match when considered together). However, in practice, we have found pair-wise matching to be sufficient. Empirical support for pair-wise matching can be found in Sec. 6.1.

A major effect of noise is that perfect matches are virtually eliminated; we cannot rely on the presence of zeros of $dist_{\mathcal{A}}$. Because $dist_{\mathcal{A}}$ is well-behaved, we can still use minima as indicators of likely periods, although it is necessary to evaluate the significance of each minimum in order to discriminate a periodic from a non-periodic motion. Towards this end, we seek a confidence measure on a real-valued function, f, with the following normalization constraints: (1) the range is [0,1], (2) a value of 1 is achieved if and only if $f(t) = 0$, and (3) the confidence is 0 if $f(t) \geq mean_f$. The following confidence formula satisfies these constraints:

$$confidence_f(t) = max(0, 1 - \frac{f(t)}{mean_f}) \tag{11}$$

This *confidence* function has one singular point, occurring when $mean_f = 0$. In this case the motion has a period of 1, a case which can easily be checked for separately.

With this machinery, we now define an algorithm for detecting if a motion is periodic and determining its period:

1. Compute $dist_{\mathcal{A}}(\mathbf{F}_i, \mathbf{F}_j)$ for all values of $i < j$ within a reasonable range (e.g., so that $\mathbf{F}_i$ and $\mathbf{F}_j$ have at least 5 features in common).

2. Average the results from Step 1 for each value of $p = j - i$; compute $M(p) = \underset{j-i=k*p}{mean}\{dist_{\mathcal{A}}(\mathbf{F}_i, \mathbf{F}_j)\}$. k varies over the integers to encourage recurrence.

3. Compute $confidence_M$ to determine if there is an acceptable period.

The above algorithm returns a list of confidence values, one for each candidate period. In practice, the detected period with the highest confidence, p_{max}, corresponds to an integral multiple of the true period. To see why this is true, notice that a periodic function will repeat at integral multiples of its period. If a confidence threshold is specified, the true period can be determined by choosing the smallest divisor of p_{max} whose confidence value exceeds the threshold.

One might hope to replace Step 3 with a Fourier transform and deduce the fundamental frequency (and hence the period) as in [4]. However, this modification is theoretically unjustified due to the fact that $M(i) \neq M(i+p)$ in general, for a periodic motion. In fact, the only constraint we use is that $M(k*p) = 0$ for positive k, which follows from the *Projected Match Criterion*.

The cost of this algorithm is dominated by the first step which calls for $O(m^2)$ two-way matches, each having a match cost of $O(n)$ operations (recall that m is the number of stills and n the total number of features). In practice the cost can be dramatically reduced by subsampling i in Step 1, without a significant decrease in accuracy[3]. Any number of values of i can be chosen based on the available computing resources. If this number is chosen to be constant, the overall cost becomes $O(mn)$. Subsampling is discussed further in Sec. 6.1.

[3]Note that subsampling does not affect the number of possible periods considered.

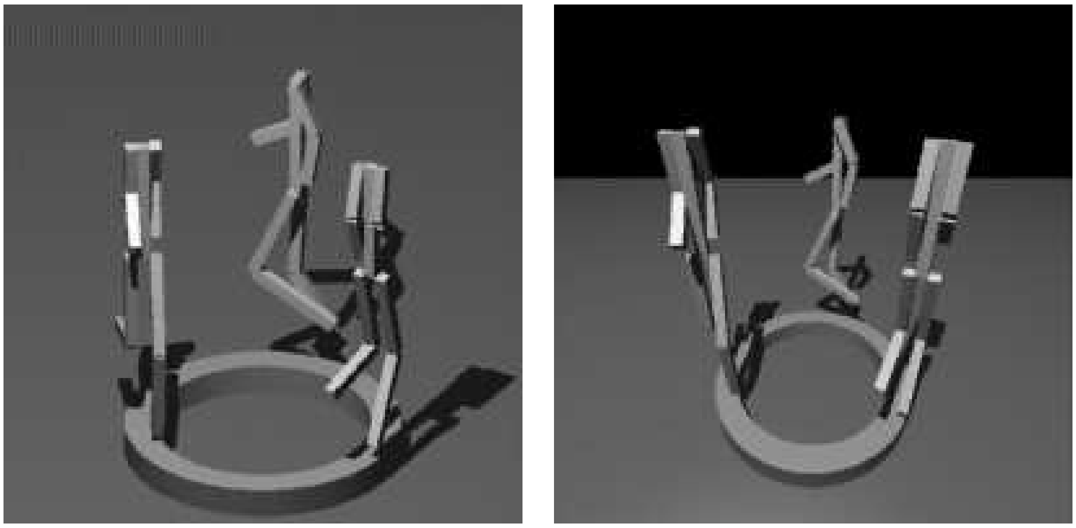

Figure 3: Projections of a simulated jogger moving around a track. Left: The jogger at three different points along the path under orthographic projection. Right: Same image under perspective projection with the camera 300 units from the center of the track. Note the projective distortions in the perspective image that are not modeled by orthographic projection.

6 Experimental Results

In this section we present the results of running the algorithm on three image-sequences, one synthetic and the other two real. The sequences are of a simulated human jogging around a track, a rotating phonograph turntable, and a person walking in an arc. The latter two sequences were taken with a moving, off-the-shelf, hand-held video camera and features were tracked using reflective markers and custom software.

6.1 Results on Synthetic Data

In order to study the effects of noise and perspective distortion on the algorithm, we created a simulation of a human jogging around a short track using a volumetric model and real motion data (see Fig. 3).

To model the human torso, we used nine parallelepipeds connected by revolute joints. The model was roughly 300 units tall and the track had a diameter of 200 units. The camera was fixed at an angle of 30 degrees from horizontal. The periodic motion consisted of a sequence of joint angles extrapolated from real motion data of a human running, provided by N. Goddard [2]. The projected positions of vertices of the parallelepipeds were used as input to the algorithm. No attempt was made to distinguish visible from occluded vertices. We found the algorithm to be robust to both feature noise and perspective effects (see Fig. 4). The period with the highest confidence was always a multiple of the correct period. The most evident effect of noise and perspective was a gradual degradation of confidence values (see Fig. 4). The results suggest that estimates of noise and perspective factors, if available, should influence threshold selection.

The algorithm in Sec. 5 computes period confidence values by matching two stills at a time. While matching a large number of stills at a time is impractical due computational expense and changes in the feature set, the choice of two stills is by no means mandatory. Fig. 5 compares confidence values for matching 2, 3, 4, and 5 stills at once. Fig. 5 shows that the confidence curves (level-curves obtained by fixing the number of stills) are qualitatively very similar. Most significantly, the true period (30) can be easily determined from any of these curves. Since the overall computational cost of the algorithm grows as the number of stills increases, it makes sense to match a small number of stills (e.g., 2 or 3) at once.

Another way of significantly reducing the run-time of the algorithm is to decrease the number of iterations (the values of i in Step 1 of the algorithm). Recall that the algorithm works by matching each still with each subsequent still. The technique can be made more efficient by choosing a small set of *reference stills* to be matched against. With this modification all matches that do not include a reference still are eliminated. A variety of methods could be used to choose reference stills, such as subsampling i (e.g., $i = 1, 3, 5, 7 \ldots$) random selection, or truncating i at some chosen point. Each method can be tailored according to available computing resources. The subsampling technique is susceptible to problems when the sampling rate and period are not relatively prime. For instance, if the sampling rate is 10 and the period is 30, the "same" three stills will be chosen in each cycle (i.e., the $1st$, $11th$ and $21st$). In this case the reference stills are redundant and it is analogous to choosing only three reference stills. Our implementation avoids this problem by using the truncation method (see Fig. 6).

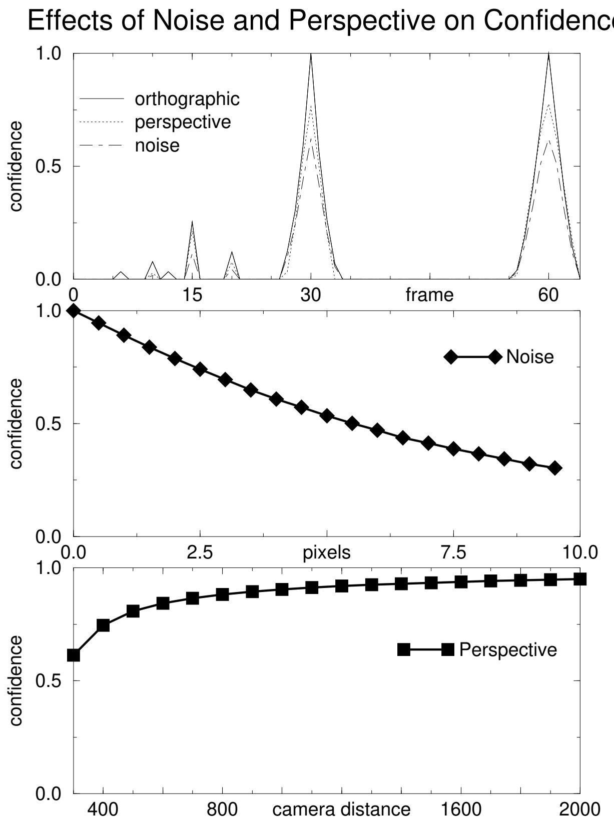

Figure 4: Effects of noise and perspective on synthetic jogger sequence. Top: Plot of period confidences for orthographic projection, perspective projection from 400 units away, and uniform feature noise in a radius of 4 pixels. Middle: Plot of confidence of the true period under varying amounts of feature noise. Each feature was perturbed randomly in a local disk of radius increasing from 0 to 10 pixels. Results were averaged over 5 trials per data point. Bottom: Confidence under perspective as camera distance from the track center varied from 300 to 2000 units. In all cases the detected period was a multiple of the true period.

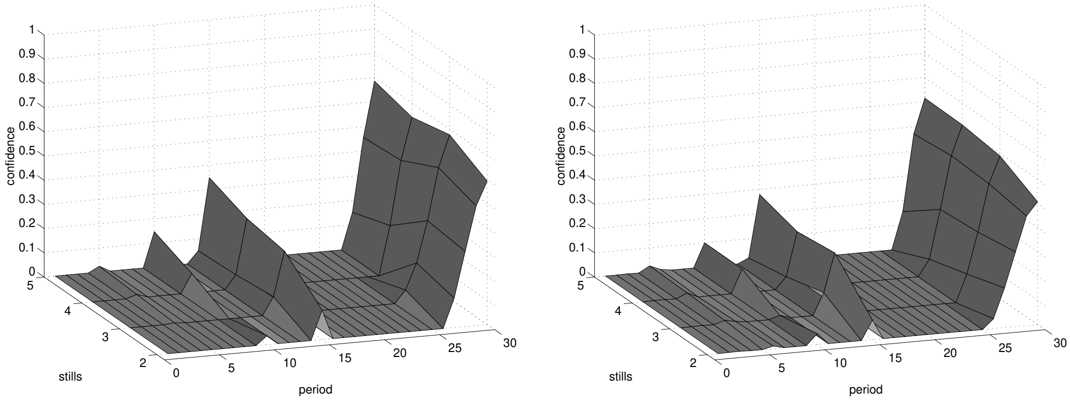

Figure 5: Effects of varying the number of stills matched on period confidence values. Left: Simulated jogger under perspective projection from 300 units away. Right: Simulated jogger under uniform feature noise of 5 pixel radius and orthographic projection. Changes in confidence values due to increasing the number of stills matched are minor whereas the increase in computational cost is significant (not shown).

Fig. 6 illustrates the use of reference stills with a synthetic sequence of a bat flapping. The motion of a bat is such that it passes through each wing configuration *twice* in each cycle, once when the wings are rising and once when they are falling. Therefore, there is a potential for "spurious" periods if too few reference stills are used. On a Sun Sparcstation 10/30 total running time was 21 seconds without the use of reference stills, less than 3 seconds with 10 reference stills, and less than $\frac{1}{2}$ second with 1 reference still. The bat sequence contained 20 features and 100 stills.

6.2 Results on Real Image Sequences

The first real image sequence contains a rotating turntable with the camera moving roughly in a quarter-arc around the turntable. Note that the algorithm cannot detect the period of an object, such as the turntable, whose only motion is affine. This limitation is easily overcome by considering both features on the moving object and elsewhere in an otherwise static scene. The algorithm detects the period of the entire scene, which corresponds to the period of the affinely-moving object. Under these circumstances, it is

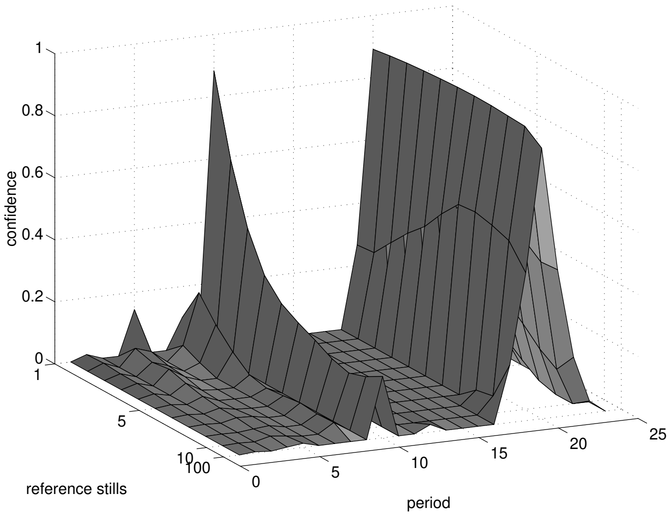

Figure 6: Effects of using a small number of reference stills on period confidence values. The algorithm was run on a synthetic flying bat sequence having a period of 20. With only one reference still a spurious period of 10 appears. As the number of reference stills is increased, the spurious period is eliminated and the confidence values quickly converge to the limit (100 reference stills is the maximum possible).

not necessary to segment the scene before applying the algorithm. As can be seen in Fig. 7, the algorithm correctly detected the true period with high confidence (0.88).

In the second image-sequence, a human subject walks in an arc subtending about 70 degrees. To aid in feature detection and tracking, reflective markers were placed in areas which were visible for the duration of the sequence (i.e., right arm, right leg, mid torso, and head). The motion was relatively even and the algorithm was able to detect a likely period of 37 with a confidence of 0.58 (see Fig. 7). The relatively low confidence value, compared to that of the turntable, can be explained by the fact that each still in the walking sequence can be roughly approximated by an affine transformation of the first still, through horizontal shears and reflections. In fact, we found this property to be true of other human locomotory motions such as running, skipping, and jumping. We have found that the periods of these nearly-affine motions can be reliably detected with our approach, but the resulting confidence values tend to be lower.

7 Conclusion

We have presented a method of detecting periodicity that is insensitive to changes in viewpoint and affine transformations of a periodically-moving object. The approach is provably-correct with ideal data and was extended to work with noisy data using a measure of confidence. We have evaluated the performance of the algorithm with respect to feature noise and perspective distortion, and have demonstrated its effectiveness on real image sequences. Future work will investigate the generalization to repeating motions lacking a constant period as well as consider the feasibility of obtaining sub-still accuracy.

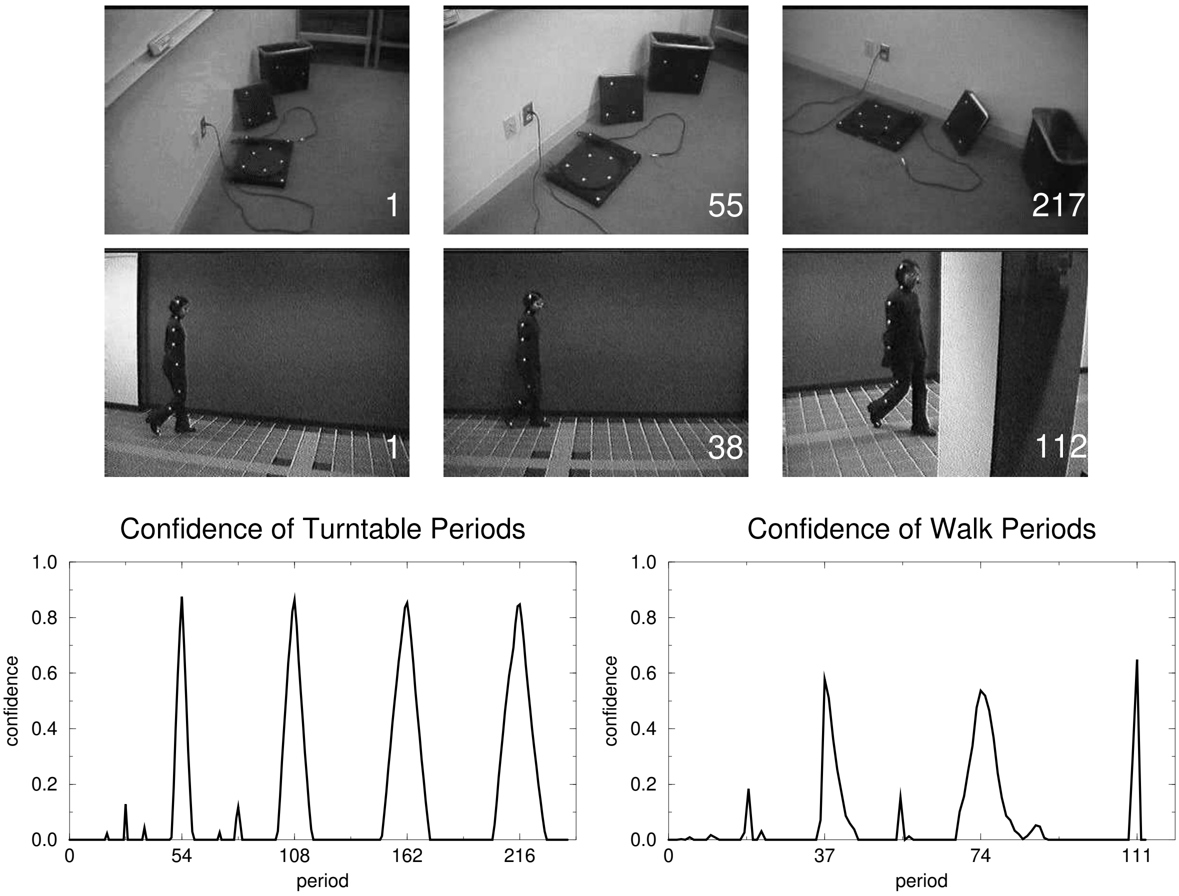

Figure 7: Results on two real image sequences. The ground truth frequency for the turntable is $33\frac{1}{3}$ revolutions per minute, or 54 stills per revolution at NTSC video rate. Despite severely uneven camera motion, the algorithm detected the true period, with a confidence of 0.88. For the sequence of a person walking, the period with the highest confidence is 111 stills. However, 37 stills is the period selected since it is an integral divisor of 111 and has a relatively high confidence value (e.g., greater than 0.5). Above are selected stills from each image sequence.

References

[1] Mark Allmen and Charles R. Dyer. Cyclic motion detection using spatiotemporal surfaces and curves. In *Proc. Int. Conf. on Pattern Recognition*, pages 365–370, 1990.

[2] Nigel H. Goddard. *The Perception of Articulated Motion: Recognizing Moving Light Displays.* PhD thesis, University of Rochester, Rochester, NY, 1992.

[3] Jan J. Koenderink and Andrea J. van Doorn. Affine structure from motion. *Opt. Soc. Am. A*, 8:377–385, 1991.

[4] Ramprasad Polana and Randal Nelson. Detecting activities. In *Proc. Computer Vision and Pattern Recognition*, pages 2–7, 1993.

[5] K. Rohr. Incremental recognition of pedestrians from image sequences. In *Proc. Computer Vision and Pattern Recognition*, pages 8–13, 1993.

[6] Steven M. Seitz and Charles R. Dyer. Affine invariant detection of periodic motion. Technical Report 1225, University of Wisconsin, Madison, WI, 1994.

[7] G. W. Stewart. *Introduction to Matrix Computations.* Academic Press, New York, NY, 1973.

[8] Carlo Tomasi and Takeo Kanade. Shape and motion from image streams under orthography: a factorization method. *Intl. Journal of Computer Vision*, 9(2):137–154, 1992.

[9] Ping-Sing Tsai and Mubarak Shah. Cyclic motion detection. Technical Report CS-TR-93-08, University of Central Florida, Orlando, FL, 1993.

[10] Daphna Weinshall. Model-based invariants for 3-d vision. *Intl. Journal of Computer Vision*, 10(1):27–42, 1993.