CYCLIC MOTION DETECTION USING SPATIOTEMPORAL SURFACES AND CURVES

by

Mark Allmen
Charles R. Dyer

Computer Sciences Technical Report #881

October 1989

Cyclic Motion Detection Using Spatiotemporal Surfaces and Curves

Mark Allmen

Charles R. Dyer

Computer Sciences Department
University of Wisconsin - Madison
1210 West Dayton Street
Madison, Wisconsin 53706

Abstract

The problem of detecting cyclic motion, while recognized by the psychophysical community, has received very little attention in the computer vision community. In this paper cyclic motion is formally defined as repeating curvature values along a path of motion. A procedure is presented for cyclic motion detection using spatiotemporal (ST) surfaces and ST-curves. The projected movement of an object generates ST-surfaces. ST-curves are detected on the ST-surfaces, providing an accurate, compact, qualitative description of the ST-surfaces. Curvature scale-space of the ST-curves is then used to detect intervals of repeating curvature values. The successful detection of cyclic motion in two data sets is presented.

The support of NSF under Grant No. IRI-8802436 is gratefully acknowledged. Code for the calculation of curvature scale-space was provided by Farzin Mokhtarian.

Table of Contents

1. Introduction 1

1.1. Properties of the Cyclic Motion Detection Problem 3

2. A Formal Definition of the Cycle Detection Problem 5

2.1. The Cycle Detection Problem for a Moving Point 6

2.2. The Cycle Detection Problem for a Rigid Object 7

2.3. The Cycle Detection Problem for an Articulated Object 9

3. Finding Cyclic Motion of Articulated Objects 10

3.1. Spatiotemporal Curves 11

3.2. Detecting Spatiotemporal Curves 13

3.2.1. Using Curvature Extrema to Place Spatiotemporal Curves 14

3.3. Finding Repeating Patterns in Spatiotemporal Curves 18

3.3.1. Curvature Scale Space 20

4. Results 23

5. Concluding Remarks 24

References 25

Appendix 1: Proof Sketch of Theorem 3.1 28

1. Introduction

Many natural objects undergo cyclic motion. Examples include a human walking, a person riding a bike, a running dog, a swinging pendulum, and a bouncing ball. The fact that people perceive these motions as repeating demonstrates that the human visual system (HVS) is capable of detecting cyclic motion. The study of humans' ability to perceive a stimulus as cyclic[1] dates back to the late 19th century [Bolt94] and has continued as an active area of research [Frai78]. While of the ability of the HVS to detect cyclic stimuli has been recognized [Frai78], there is little research in the area. Most of the research concentrates on the human perceptual system's (HPS) ability to perceive auditory rhythm. In this paper we will examine the detection of cyclic stimuli in vision.

The human perceptual system's ability to perceive rhythm can be divided into three categories: rhythmical grouping, forming strings of rhythmical groups, and experience of rhythm [Frai78]. Rhythmical grouping is demonstrated by the fact that a series of identical sounds is spontaneously perceived as groupings of two, three or four elements [Bolt94]. A similar result has been obtained using light stimuli [Koff09]. Fraisse [Frai75] showed that rhythmical groups are gestalt, meaning that each group constitutes a functional unit. These units are related to each other by temporal order and pauses to form strings of rhythmical patterns. For example, a rhythmical group follows another after a pause. When these patterns are then linked to what follows, rhythm is experienced. Clearly, the perception of rhythm can be accompanied by a recognition of the rhythm. For example, when Gabrielsson [Gabr73] presented subjects with complete musical patterns and measured their perception of rhythm, the subjects could also recognize the pattern if they had heard it before. The analog of this in the visual domain was shown by Johansson [Joha73] using moving light displays (MLD). While the recognition of the rhythms was not central to those studies, one can extrapolate from the results and conclude that the perception of rhythm is accompanied by recognition. This paper addresses the detection of cyclic motion within the context of motion recognition.

There has been considerable work examining the HVS's ability to recognize movement that contains cyclic motion, but the detection of the cycles was not the focus of the studies [Joha73].

---

[1]The psychology community refers to a repeating stimulus as "rhythmic" rather than "cyclic". The two are used interchangeably in this paper.

Johansson's MLD's contained cyclic motion in the form of human walking, running, cycling and dancing. When subjects were presented with an MLD of walking, they always recognized the motion after the first one or two steps.

The cyclic motion of each joint in a walking MLD is clear in Figure 1.1. In addition to each joint undergoing cyclic motion, the entire body is undergoing cyclic motion. Consider, for example, the motion of the wrist from the time when the left hand is at the back of its swing until the time that the left hand is again at the back of its swing. This constitutes one period of cyclic motion of the body and is easily recognized when viewed as an MLD as well as when viewed in real, natural scenes.

Unfortunately, Johansson's work examined subjects' ability to recognize the high-level motion of an MLD, not any particular cyclic motion contained in the MLD. In this paper we address the problem of recovering a cyclic motion description from an input sequence using computational methods. This problem, while recognized by the psychology community, has been largely ignored by the computer vision community. By addressing the computational issues we can better understand the HVS's ability to perform this task.

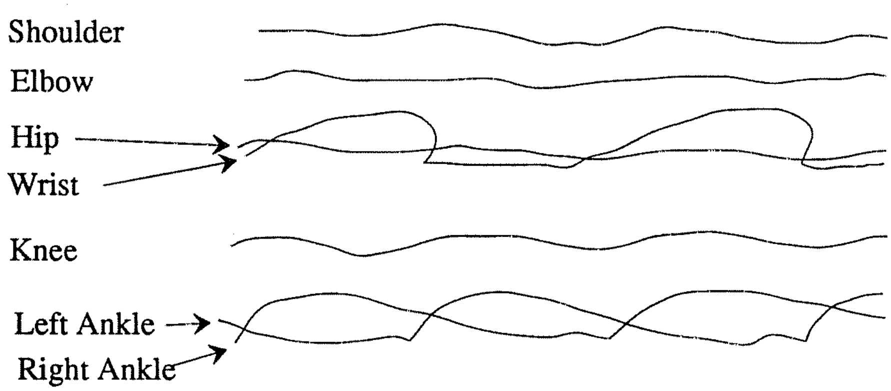


Figure 1.1. Typical motion paths of seven elements representing the motions of the right side joints plus the ankle joint of the left leg of a walking person. [Joha73]

1.1. Properties of the Cyclic Motion Detection Problem

By examining the HVS's ability to recover and describe cyclic motion, we can learn pertinent properties of the cyclic motion detection problem. Further, these properties give a better understanding of the problem and suggest a form of a solution. These properties include: 1) recognition of objects is not necessary for the recovery of cyclic motion; 2) many frames are necessary before cyclic motion becomes evident; 3) cyclic motion does not depend on absolute position; and 4) cyclic motion can occur at multiple scales.

Motion can be described at many levels, from low-level, detecting that something in the scene has moved between two frames, to high-level, recognition of a "coordinated sequence of events" [Godd88]. Most previous work has focused on low-level motion analysis. Frame differencing of temporally-adjacent frames is one common method of low-level motion analysis [Jain79, Jain79a, Jain81, Jain83, Jaya83, Ande85, Lee88].

High-level motion analysis, i.e., recognizing a coordinated sequence of events such as walking and throwing, has been formulated previously as a process which follows high-level object recognition [Hogg83, Akit84, Leun87, Leun87a]. Only recently have some researchers considered high-level motion analysis as a process which does not depend on complex object descriptions [Godd88, Godd88a, Godd89, Goul89].

In order to substantiate the hypothesis that motion recognition can occur before object recognition, we will show in this paper that detecting cyclic motion in an image sequence can be performed without prior recognition of the object undergoing the cyclic motion. Other results also indicate that the perception of intermediate descriptions is accomplished prior to recognition of the stimulus. For example, Barrow and Tenenbaum [Barr81] used photomicrographs of pollen grains to argue that the HVS can recover surface shape and orientation without recognition. When Gabrielsson [Gabr73] presented subjects with songs, they could detect rhythm within the music even without having heard the song previously. These psychophysical results suggest that intermediate-level descriptions can be recovered prior to object recognition. Further, it is reasonable to conclude that if the HVS is presented with a "movie" of photomicrographs of pollen grains moving in a cyclic fashion, that the HVS could detect the cyclic motion even though it has no familiarity with the objects.

Since cyclic motion of non-recognizable objects, e.g., pollen grains, can be recovered, static models of objects are not a necessary component of a solution. Rather than a model-based solution, a motion-based solution is required. It is not the object itself that is the appropriate level of motion description, but rather the motion of lower-level primitives such as surfaces.

Cyclic motion is an important problem because it is a general example of an intermediate-level motion description. It is a lower-level description than walking, for example, but it requires more than simply making changes between a few frames explicit, as in low-level motion analysis. Since cyclic motion is an intermediate-level description, one would expect that it should be computed from a low-level motion description. The fact that even the simplest of cyclic motions, e.g., a swinging pendulum, requires many frames before its cyclic behavior is recoverable, suggests that the low-level motion representation must describe long-range motion sequences.

Recently, there has been a trend toward using more than a few frames to analyze an image sequence [Jain88, Aloi88]. Cyclic motion description requires long sequences of frames since many frames are required before a motion repeats. These considerations suggest a low-level representation that coherently represents many frames. Frame differencing is insufficient since only the differences between two frames is made explicit. We choose to use dense spatiotemporal image sequences, created by "stacking" many frames together, as the low-level motion representation. The concise, coherent nature of this representation has also lead others to use it to examine long sequences of frames [Boll87, Bake88, Aloi88, Liou89].

Just as cyclic motion detection does not depend on the object undergoing the cyclic motion, as we argued earlier, it also does not depend on the absolute position where the motion takes place. It is not the position that is essential to characterizing cyclic motion, for example, but rather how the object moves that is important. For example, the motion produced by a person walking parallel to the image plane would be called cyclic. This perception of cyclic motion is not dependent on the absolute position of the person because the person is never in the same place at any two separate times. So the definition and detection of cyclic motion must be invariant to position.

Finally, cyclic motion can occur at multiple scales. For example, consider a bird flying in a cyclic pattern. The bird displays the cyclic motion of its flight pattern and also the finer cyclic motion of its flapping wings. Thus, a complete cyclic motion description should describe all levels of cyclic motion.

In summary, we have argued that cyclic motion detection and description

- does not depend on prior recognition of objects

- does not depend on absolute position information
- must make use of long-range temporal sequences
- must be sensitive to multiple scales

The rest of this paper is organized as follows. In Section 2 cyclic motion is formally defined as repeating curvature along a path of movement. Spatiotemporal (ST) surfaces are introduced in Section 3 as the appropriate low-level motion description from which cyclic motion is recovered. Section 3 continues by showing that cyclic motion in the scene which is preserved under projection is retained by the ST-surfaces. In Section 4 it is shown how to recover the cyclic shape of ST-surfaces using ST-curves and curvature scale-space.

2. A Formal Definition of the Cycle Detection Problem

Before we can define cyclic motion we must have a representation of motion. We need a representation that makes the essence of cyclic motion explicit. One possibility is to represent the motion of an object by its position over time. However, we are more concerned with *what* kind of motion occurs rather than *where* it occurs [Rubi85]. Thus, it is not position that is essential to characterizing cyclic motion, but rather how the object moves that is important. For example, the motion produced by a person walking parallel to the image plane is called cyclic. But clearly this perception of cyclic motion is not dependent upon the position of the walking person because they are never in the same place at any two separate times. We need a definition of cyclic motion that is invariant to the position of the object. Curvature and torsion along the path of movement of an object represents movement and is invariant to position. In fact, the curvature and torsion along the path uniquely defines the path up to a rigid transformation [DoCa76]. However, for our purposes, curvature alone will be sufficient to describe how an object moves. We choose this over torsion because it is possible that some object's motion is restricted to a plane, e.g., a swinging arm.. The torsion along this path of motion would always be 0 since torsion along a planar curve is always 0. Curvature, on the other hand, will not always be 0.

In the remainder of this section we define the cycle detection problem for a single point and then extend it to more complicated objects. Having defined cyclic motion, we will apply the definition to spatiotemporal surfaces and curves. We will prove that if an object undergoes 3D cyclic motion which is preserved under projection, then the resulting spatiotemporal surfaces must retain this cyclic information. We will show how to place spatiotemporal curves on spatiotemporal surfaces and how to use the curves to recover the cyclic information.

2.1. The Cycle Detection Problem for a Moving Point

Consider the situation of a single point moving in three dimensions. Let

$$\alpha\colon \mathrm{T} \rightarrow \mathbb{R}^3 \qquad\qquad \alpha(t) = (x,y,t)$$

define the path of a point over the temporal interval T. Since α defines a path in 3-space it is a space curve and therefore properties associated with spaces curves, such as curvature, can be applied to the curve defined by α. We make no assumptions about the smoothness of α.

Let

$$\kappa(t) = \frac{\sqrt{A^2+B^2+C^2}}{(\dot{x}_1^2 + \dot{y}_1^2 + \dot{t}_1^2)^{3/2}}$$

where

$$A = \begin{vmatrix} \dot{y}_1 & \dot{t}_1 \\ \ddot{y}_1 & \ddot{t}_1 \end{vmatrix} \qquad B = \begin{vmatrix} \dot{t}_1 & \dot{x}_1 \\ \ddot{t}_1 & \ddot{x}_1 \end{vmatrix} \qquad C = \begin{vmatrix} \dot{x}_1 & \dot{y}_1 \\ \ddot{x}_1 & \ddot{y}_1 \end{vmatrix}$$

define the curvature at $\alpha(t)$. Since we made no assumptions about the smoothness of α it is possible that for some t, $\kappa(t)$ is undefined. In this case we will say that $\kappa(t) = \infty$. And for t_1 and t_2, if $\kappa(t_1) = \infty$ and $\kappa(t_2) = \infty$ then $\kappa(t_1) - \kappa(t_2) = 0$.

We say that α undergoes *Continuous Cyclic Motion* if:

$$\max\left\{ |\,\kappa(t) - \kappa(t + (t_2 - t_1))\,| \;:\; t \in [t_1,t_2] \right\} < \varepsilon$$

With this definition, only two periods of cyclic motion are needed before it can be classified as cyclic.

This definition refers to motion in the 3D scene. When we make use of this definition in later sections, we will refer to 3D cyclic motion that is preserved under projection, i.e., cyclic motion along the axes perpendicular to the line of sight. Nature does not usually conspire against us so we assume that if there exists cyclic motion in the image sequence, then there existed cyclic motion in the scene.

There are a few points worth making about the types of movement that would be considered cyclic under this definition. The definition *looks* for time intervals during which the curvature in the intervals is equal. Consider a point moving in a straight line. The curvature is always zero. So given any two intervals, the curvature will always be equal during these intervals. In this degenerate case, a point displays cyclic motion with an arbitrary period.

Consider a point moving in a circle with constant speed. This is equivalent to the previous case in that the point displays cyclic motion with an arbitrary period. But it also fails to distinguish the cycle where one period equals one revolution of the circle. Perhaps this is undesirable, but in order to detect one revolution of the circle as one period the definition would have to be sensitive to the position of the point over time which, as argued earlier, is undesirable in most cases.

Goddard [Godd88, Godd88a, Godd89] used a change in angular momentum as the primitive to detect movements such as walking. A point moving in a circle with constant speed does not change its angular momentum, so it would also be missed using Goddard's definition. Similarly, the Trajectory Primal Sketch of Gould and Shah [Goul89] is not sensitive to such cases. A definition of cyclic motion that is sensitive to such cases would only obscure the type of motion of the point. As stated earlier, it is the type of motion that we wish to detect.

2.2. The Cycle Detection Problem for a Rigid Object

Given our definition of cyclic motion for a point, we now extend the definition to rigid objects. This extension is motivated by the following theorem:

Theorem 2.1: If a point of a rigid body displays cyclic motion, then with probability 1 all other points on the rigid body display cyclic motion with the same period.

Proof Sketch: Any instantaneous motion of a rigid body in three dimensions can be uniquely described by a *twist* [Coxe61]. A twist is described by an axis l, an angular velocity ω, and a translational velocity v along l. Since a twist describes the motion of any rigid object, it can be applied to an entire rigid object or a point on the rigid object. In other words, a point on a rigid object can be thought of as a rigid object. Therefore, the point's motion can be defined by a twist. Since a twist is unique, the twist describing the motion of the point must be the same as the twist describing the motion of the entire rigid object that contains the point. Let p_1 be the point feature of the solid object that displays cyclic motion. Let $twist_t$, defined by l_t, ω_t and v_t describe the motion for p_1 at time t. The twists along the path of p_1 determine the curvature of the path of p_1. Let $\kappa(p_1, t)$ define the curvature of p_1 at time t. In the simplified case where v remains zero, $\kappa(p_1, t)$ equals $r_{p_1,t}$ where $r_{p_1,t}$ is the perpendicular distance from p_1 to l_t (Figure 2.1). In the case where v is not 0, the curvature along the path of movement of a point is a function of l, ω, v and their first and second partial derivatives with respect to arc length along the path of movement.

We need to show that given the curvature of p_1 at t, the curvature of another point p_2 at t is dependent upon the curvature of p_1 at t. That is,

$$\kappa(p_2,t) = \delta\,\kappa(p_1,t)$$

for some δ. From Figure 2.1 it is clear that in the simplified case where $v_t = 0$, $\delta = \frac{r_{p_2,t}}{r_{p_1,t}}$. In the case where $v_t \neq 0$, δ is a function of $\frac{r_{p_2,t}}{r_{p_1,t}}$, v and their partial derivatives. This shows that given

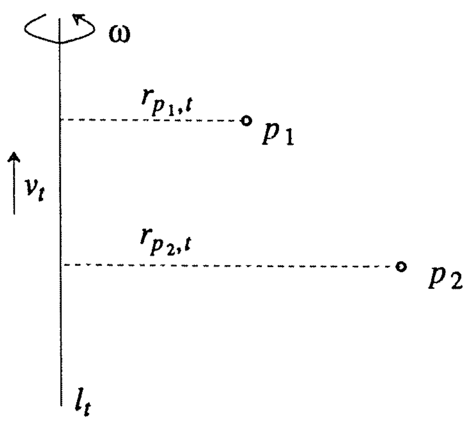


Figure 2.1. A *twist* at time t.

the curvature of p_1 at time t we can find the curvature of any other point at time t.

p_1 is the point undergoing cyclic motion. So the sequence of curvature values for p_1 is repeating. The sequence of curvature values for p_2 is determined by the sequence of curvature values for p_1. So if p_1 is cyclic, p_2 is also cyclic. This assumes that $r_{p_1,t}$ is never 0 so δ is always defined. In the continuous case, the probability that p_1 lies exactly on l_t is 0. Therefore, with probability 1, if a point on a rigid object undergoes cyclic motion, then all other points on the rigid object undergo cyclic motion with the same period.

□

This proof assumes the objects are moving in 3-space. One would like to know the effect of the projection into a 2D image. In this initial study we assume objects do not rotate in depth[2]. So if a point projects into the image in one frame, it will always project into the image. If a point undergoes cyclic motion, its projection will also undergo cyclic motion, as will all other points of the object that project into the image.

If any point on a rigid object undergoes cyclic motion, all other points on the rigid object undergo cyclic motion with the same period. So we say a rigid object undergoes cyclic motion if any point feature of the object undergoes cyclic motion. And a point undergoes cyclic motion as defined in Section 2.1.

2.3. The Cycle Detection Problem for an Articulated Object

An articulated object is composed of rigid objects connected by joints. A rigid part of an articulated object undergoes cyclic motion as defined above. We need a definition of cyclic motion for a set of rigid objects connected by joints. A set of rigid parts of an articulated object undergoes cyclic motion if each rigid part undergoes cyclic motion and there exists a dependency between the periods of cyclic motion. A dependency exists between solid parts if the ratio of their periods remains constant. For example, the ratio of the period of a swinging forearm and the period of the upper arm remains constant for a walking person.

This definition is motivated by the fact that most articulated objects that the human visual system would classify as displaying cyclic motion have this dependency property. The definition

---

[2]This assumption will be dropped in future work (see Section 5).

prevents jointed, rigid objects with no consistent pattern to the periods of their parts from being classified as displaying cyclic motion.

3. Finding Cyclic Motion of Articulated Objects

We argued earlier for the need of a low-level motion representation that can describe long-range motion sequences. Spatiotemporal (ST) volumes are a low-level motion representation that can represent the motion of objects for arbitrary length sequences. These volumes are three-dimensional structures that are built by "stacking" a dense sequence of image frames. The ST-volume is a viewer-centered, three-dimensional (x-y-time) description. If we sample densely enough in time and an edge operator is applied [Boll87], the ST-volume will contain surfaces and volumes created by the surfaces. These surfaces and volumes represent object motion swept out through time. ST-surfaces will be used to detect cyclic motion in the scene. This is made possible by the following theorem:

Theorem 3.1: If a solid object in a scene displays cyclic motion such that the cyclic motion is preserved under projection, the spatiotemporal surfaces that correspond to intervals of cyclic motion will be piecewise isomorphic.

Proof Sketch: Consider a spatiotemporal surface such that there exist ridges on the surface and every point on a ridge has the same value for the time coordinate. We call this ridge a *temporal ridge* since it results from a nonsmooth change of velocity of an object at some time. Define a *patch* of an ST-surface as the surface between two temporal ridges. Two surfaces are piecewise isomorphic if the first patch of surface 1 is isomorphic to the first patch of surface 2, the second patches are isomorphic, etc. A curve, C, in a sequence of frames generates an ST-surface. The type of movement, translation, rotation or a combination of both, determines the surface generated by C. Given a parameterization of the ST-surface in terms of C and the movement of C, an isomorphic mapping between the corresponding patches of the ST-surfaces can be defined. Details of this proof are given in Appendix 1.

□

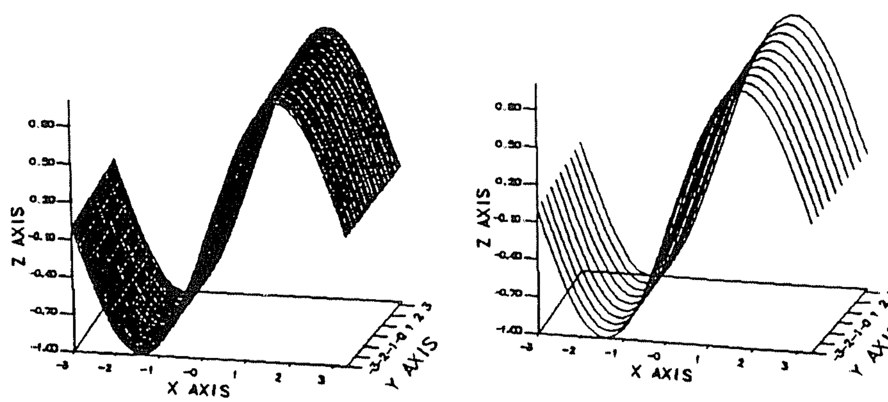
While ST-surfaces retain cyclic motion in an image sequence, methods are still required to recover the cyclic behavior of the ST-surfaces. ST-curves, defined in the next subsection, are used to recover the cyclic behavior of ST-surfaces. ST-curves are detected such that the curvature along the curve is cyclic whenever the ST-surfaces are cyclic. Once the ST-curves are detected on the ST-surfaces, cyclic curvature along the curves is found using curvature scale-space.

3.1. Spatiotemporal Curves

In this section we define the term *spatiotemporal curve* (ST-curve). We also show how to detect ST-curves on spatiotemporal surfaces so that a set of ST-curves describes the surface.

Figure 3.1a shows a sinusoidal surface. Figure 3.1b shows the same surface represented by a set of ST-curves. ST-curves are space curves and therefore properties such as curvature can be applied. ST-curves will be detected on ST-surfaces such that a repeating pattern in an ST-curve implies the ST-surface is cyclic. In Section 3.2 we discuss in depth how to detect ST-curves so that they accurately represent the surface.

An ST-curve is any curve on a spatiotemporal surface such that the curve moves forward in time, i.e., the time component of the curve is strictly increasing. Consider p_1 and p_2, two points on a spatiotemporal surface (Figure 3.2). Let S_1 and S_2 be two congruent surface patches that contain p_1 and p_2, respectively. That is, S_1 and S_2 are neighborhoods of p_1 and p_2. Since S_1 and



(a) (b)

Figure 3.1. A surface and its representation with curves.

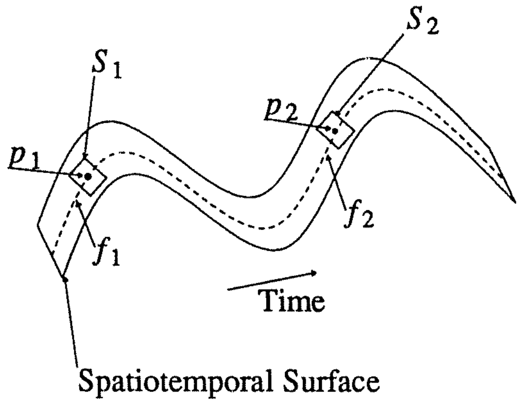


Figure 3.2. f_1 and f_2 are consistent since the tangent vectors of f_1 and f_2 at p_1 and p_2 are congruent.

S_2 are congruent, there exists some translation, T, and/or rotation, R, which when applied to S_1 makes S_1 coincide with S_2. Given an ST-curve f_1 containing p_1 and an ST-curve f_2 containing p_2, f_1 and f_2 are *inconsistent* if there exists some neighborhoods of p_1 and p_2, S_1 and S_2, such that S_1 and S_2 are congruent but the tangent vectors of f_1 and f_2 at p_1 and p_2 do not coincide after translation T and rotation R. If f_1 and f_2 are never inconsistent, they are called *consistent*. In other words, two ST-curves are consistent if they go in the same relative direction for any two congruent neighborhoods of points on the ST-curves. Note that f_1 and f_2 can be different intervals of the same ST-curve.

Theorem 3.2: If an ST-curve on a spatiotemporal surface is consistent and the surface represents cyclic behavior, i.e., the surface is piecewise isomorphic, then the curvature of the ST-curve will be cyclic.

Proof: If we can show that the curvature at every point of ST-curve f_1 within surface patch S_1 equals the curvature at the corresponding point of ST-curve f_2 within congruent surface patch S_2, f_1 must preserve any cyclic behavior present in the surface. So we must show that given a translation T and rotation R that makes S_1 and S_2 coincide, applying T and R to f_1 results in f_1 having the same curvature as f_2 at every point.

Since the ST-curve is consistent, corresponding tangent vectors along the curve coincide after a translation and rotation. Since the tangent vectors along f_1 and f_2 coincide, we are left to show that rotation and translation of all the tangent vectors by T and R does not affect the curvature of the space curve. This is equivalent to showing that translation and rotation of a space curve does affect its curvature. This is a well know fact [DoCa76].

Since the curvatures of f_1 and f_2 are always equal for corresponding points, the ST-curves will be cyclic whenever the ST-surface is cyclic.

□

All consistent ST-curves that lie on an ST-surface generated by the projection of a rigid part will be similar. For example, consider an arm swinging in a scene such that the swinging motion is preserved under projection. The projection of the swinging arm will generate an ST-surface. All consistent ST-curves detected on this surface will be similar. In particular, Theorems 2.1 and 3.2 show that the ST-curves on this ST-surface will all be cyclic with the same period. Using this property we can cluster ST-curves based on their cyclic behavior. In the current implementation, we assume that if a group of ST-curves is cyclic with the same period, then the curves lie on an ST-surface generated by the projection of a single rigid object or a rigid part of an articulated object.

3.2. Detecting Spatiotemporal Curves

From Theorem 3.2, our only requirements for detecting ST-curves are that they move forward in time and they are consistent. Currently, we also assume that no objects in the scene undergo rotation in depth. This assumption was made only to simplify the procedure for tracking points and is not a fundamental limitation of the method. Section 5 describes a method of detecting ST-curves that does not require the assumption of no rotation in depth.

We now propose two methods for detecting ST-curves on ST-surfaces: (1) using curvature extrema, and (2) using principle curvature directions. Both methods satisfy our requirements that ST-curves move forward in time and are consistent. The description of method 2 is briefly outlined in Section 5.

3.2.1. Using Curvature Extrema to Place Spatiotemporal Curves

If we take a temporal slice at time 0, we have the edge map of the image of a scene at that time. These edge points can be connected into contours and the curvature extrema detected. Doing this for each time slice and connecting corresponding curvature extrema over time defines a set of ST-curves. Since curvature extrema are relatively sparse and we assume a dense sampling of images (no more than a few pixel movement between frames), connecting the corresponding curvature extrema is reasonable.

ST-curves associated with curvature extrema must be consistent. Since the object is rigid, an ST-curve that is defined by correctly tracking a curvature extremum must correspond to a unique point feature on the object. Therefore, when the ST-surface repeats because of cyclic motion, a curvature extremum ST-curve will be repeating since it's tracking a single point of the object. This results in the ST-curve being consistent.

To experimentally test our method for detecting ST-curves, curvature extrema were tracked in two sets of data. The first set of data consisted of 66 frames of a cube translating parallel to the image plane. The first ten frames are shown in Figure 3.3. The second test data set consisted of 33 frames of a single object with one joint and its two major parts undergoing a "flapping" motion. The two parts "flap" with different periods. The first ten frames are shown in Figure 3.4. The top half of the resulting spatiotemporal surfaces for the translating box and the flapping wings are shown in Figures 3.5 and 3.6, respectively.

Test images were input as 128 by 128 binary images created by a Difference-of-Gaussian edge operator. The outer contour of the first frame was computed and the curvature at each point was calculated.

All curvature extrema, positive and negative, whose absolute value was at least three times greater than the average of the absolute value of the curvature at all points were detected in each frame. The curvature extrema in the first image were each linked to their nearest curvature extremum in the next frame. The curvature extrema in the second frame were linked to the curvature extrema in the third frame, etc. The linking of the extrema into ST-curves is shown in Figures 3.7 and 3.8.

The ST-curve in the middle of Figure 3.8 started out following the curvature extremum formed by the two rigid wings. But as the wings move this curvature extremum disappeared and then reappeared. As can be seen, it failed to correctly follow this extremum from start to finish.

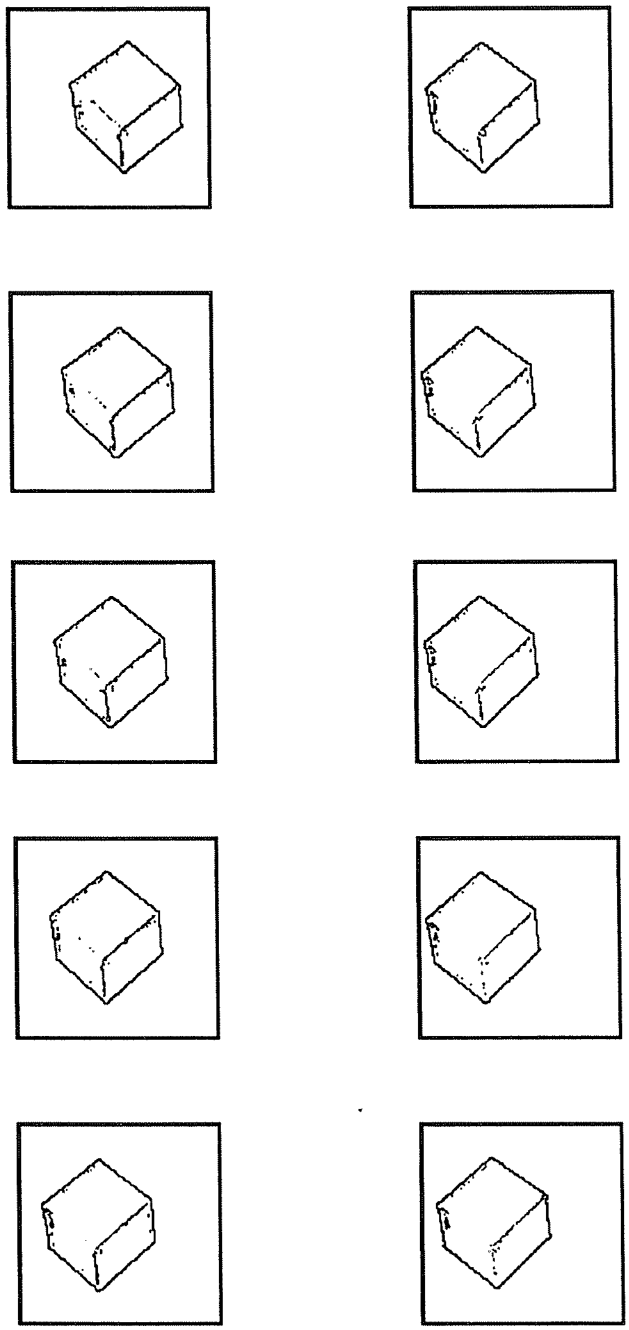


Figure 3.3. Ten frames of a translating box image sequence.

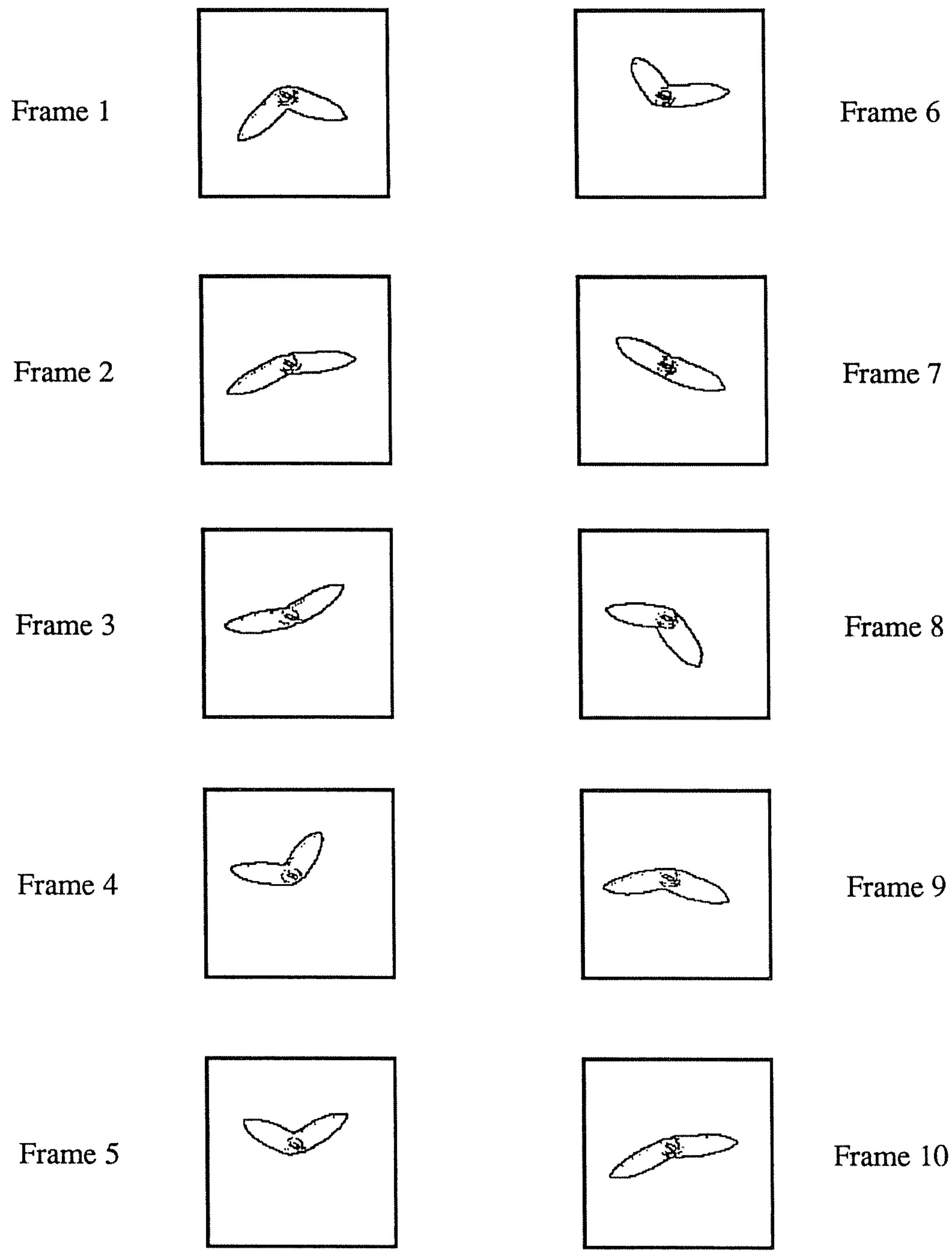


Figure 3.4. Ten frames of a flapping wings image sequence.

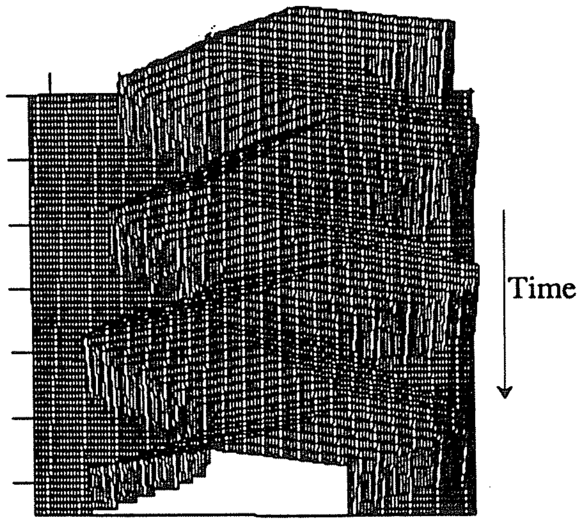


Figure 3.5. Top half of the spatiotemporal surface for the translating box image sequence.

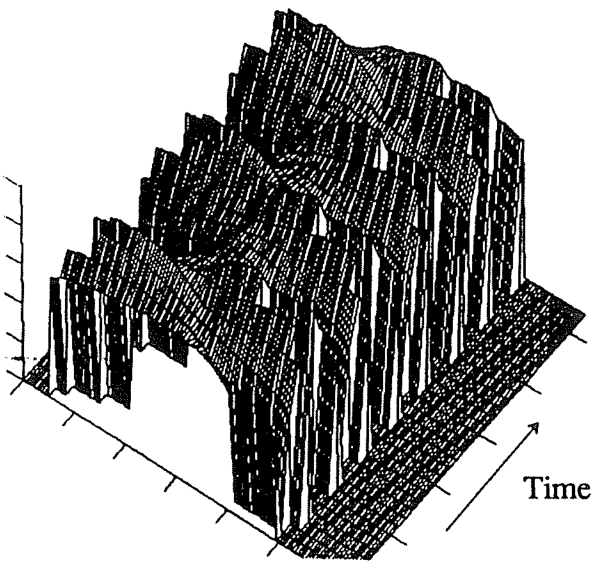


Figure 3.6. Top half of the spatiotemporal surface for the flapping wings image sequence.

This disappearance and appearance of curvature extrema is a typical occurrence with articulated objects. It may be possible, however, to define a grammar, analogous to Leyton's [Leyt88], in order to better formalize these kinds of events.

3.3. Finding Repeating Patterns in Spatiotemporal Curves

Given a set of consistent ST-curves extracted from ST-surfaces, we can now infer cyclic motion from cyclic structure in the ST-curves. All ST-curves which lie on an ST-surface generated by the projection of a rigid object, or rigid part of an articulated object, will have the same period of cyclic motion. ST-curves that track other point features, such as surface markings or vertices, will also have the same period of cyclic motion. We propose using curvature scale-space to detect repeating patterns in ST-curves.

Curvature scale-space is used to detect repeating patterns in ST-curves for many reasons. Curvature scale-space has many properties that make it desirable, regardless of the application. These are discussed below. One property, however, is particularly relevant to the problem of cyclic motion detection. Cyclic motion can occur at many scales. Since curvature scale-space represents curvature over many scales, it is a natural representation to use. Fine cyclic motion, e.g., the flapping wings of a bird, will be observable at fine scales in the scale-space, whereas coarse cyclic motion, e.g., a cyclic flight path of a bird, will appear at coarse scales in the scale-space.

By looking for repeating patterns in ST-curves, we are effectively matching portions of space curves. Mokhtarian [Mokh88] used curvature and torsion scale-space for model-based matching. Curvature and torsion scale-space can be constructed efficiently and they are sensitive to small changes in a curve. Curvature and torsion scale-space images represent information at multiple levels of detail. And scale-space images of curvature and torsion are robust since the effect of missing data at end points has only a local effect on the scale-space. This is important since we have to match portions of the scale-space generated from the middle of a space curve to portions generated at the ends of space curves.

Yuille and Poggio [Yuil83] showed that torsion scale-space is unique by showing that almost all curves can be reconstructed up to an equivalence class from the torsion scale-space. The same has not been be proven for curvature scale-space, but curvature scale-space is rich enough in detail that it is unlikely that curves with different shapes will give rise to similar curvature scale-space images [Mokh88].

Finally, curvature and torsion scale-space is position invariant. When looking for repeating patterns we do not want to concern ourselves with spatial information; in most every case spatial

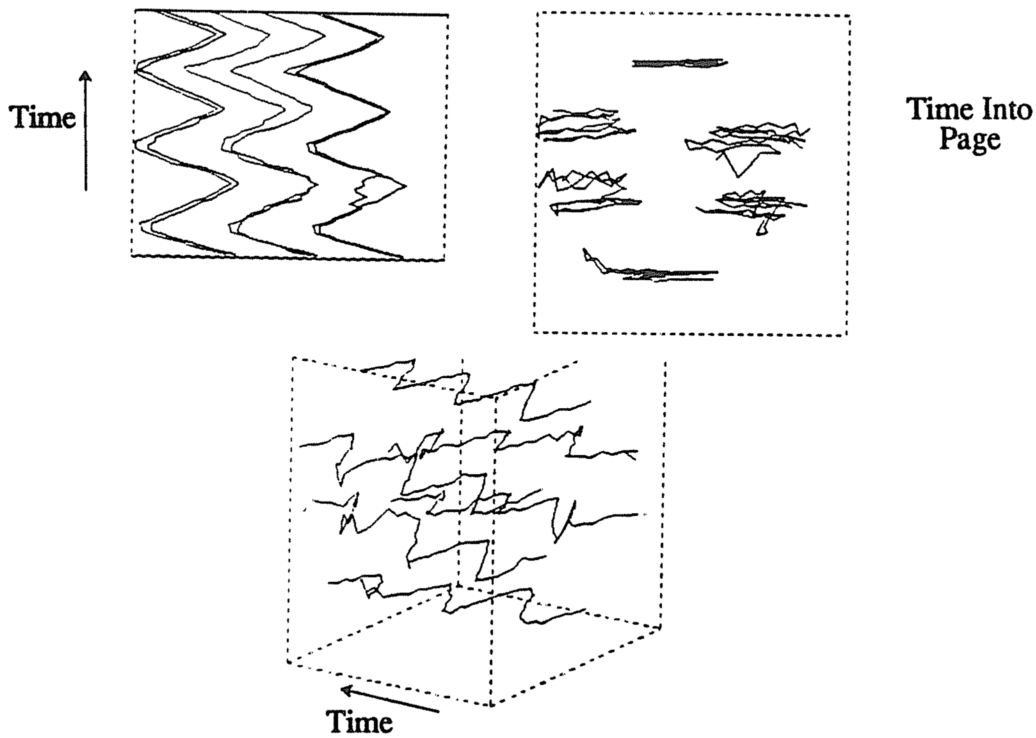


Figure 3.7. Spatiotemporal curves for the translating box image sequence.

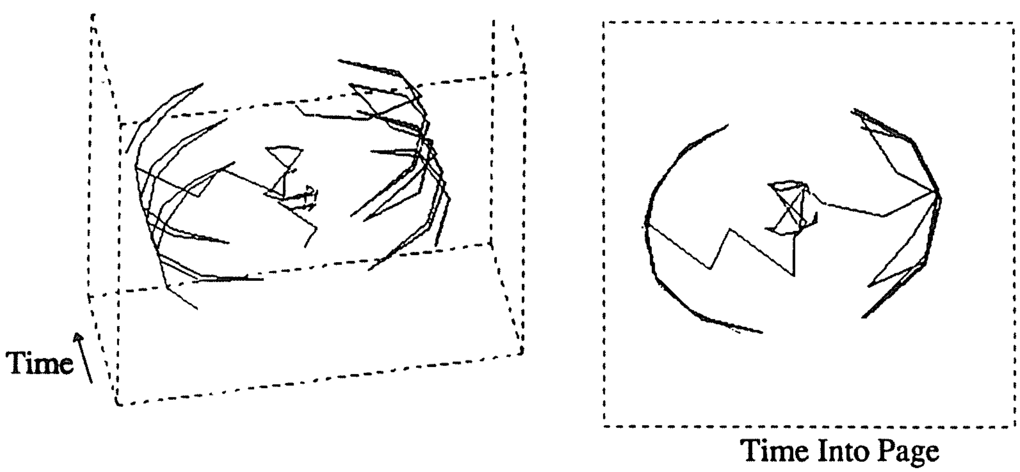


Figure 3.8. Spatiotemporal curves for the flapping wings image sequence.

information is not relevant. For example, when considering the repeating motion of a swinging arm while walking, the position of the arm over time is not what is important because it is constantly moving in the direction of walking; rather it is the series of movements that the arm makes that is important. It is known that the curvature of an ST-curve is invariant to translation and rotation [DoCa76] so the curvature scale-space is also invariant to translation and rotation. The same can be shown for torsion of an ST-curve.

3.3.1. Curvature Scale Space

Mokhtarian [Mokh88] showed how to construct curvature and torsion scale-space images for a space curve. For portions of the curve that are congruent, we expect the corresponding portions of the scale-space to be similar since curvature scale-space is probably unique. Since scale-space represents large features as well as small ones, we can use a coarse-to-fine search procedure, first looking for repeating patterns at coarse scales where the matching is less computationally demanding, and then moving to finer scales.

Figure 3.9 shows several curvature scale-space images. The scale-space image of a 1D signal is guaranteed to be composed of "arches" [Witk87]. The features in the curvature scale-space images in Figure 3.9 resemble arches. To follow the similarity with the features in a 1D scale-space image, we define a curvature scale-space feature to be an arch. This is consistent with Mokhtarian's definition of a feature [Mokh86, Mokh88].

Figure 3.9 shows features found in the curvature scale-space for the ST-curves in the translating box image sequence. Figure 3.10 shows the features found in the curvature scale-space for the ST-curves in the flapping wings image sequence.

To detect repeating patterns in the features, we used a uniform cost algorithm that is similar to Mokhtarian's [Mokh86]. The final result of the algorithm is a labeling of the features in curvature scale-space that constitutes the best match of the repeating patterns in a given curvature scale-space. The algorithm begins by creating a node for every possible pair of features in scale-space. A match cost is computed for each node, measuring how *different* the two features are. The higher the cost, the more dissimilar the features are. Next, the node with the lowest cost is expanded. The cost of expansion is added to the node's cost. Expansion of the lowest cost node continues until a node reaches a solution, i.e. the repeating pattern has been verified over the entire domain.

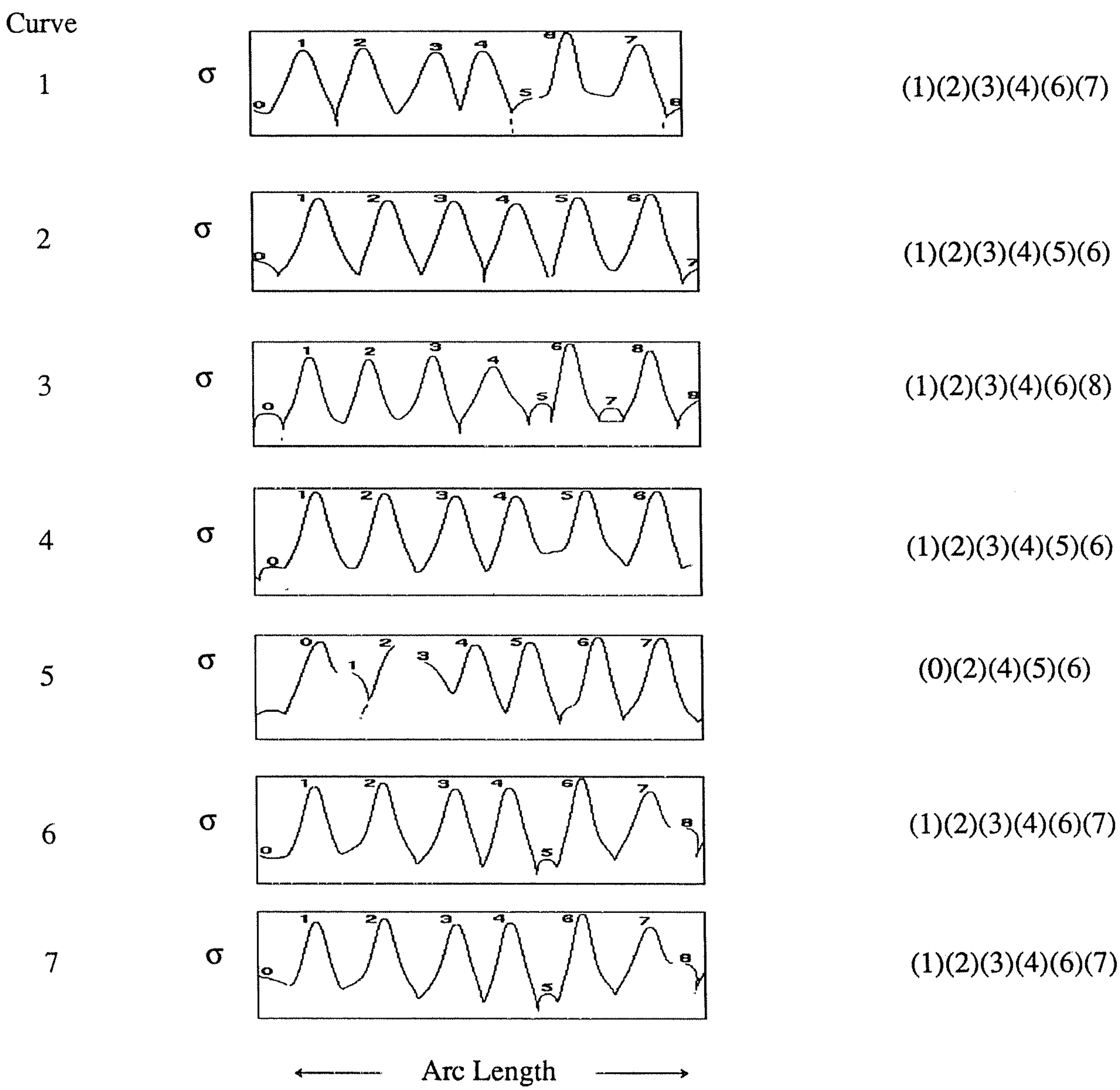


Figure 3.9. Curvature scale spaces images for the spatiotemporal curves of the translating box image sequence. The repeating patterns are shown on the right.

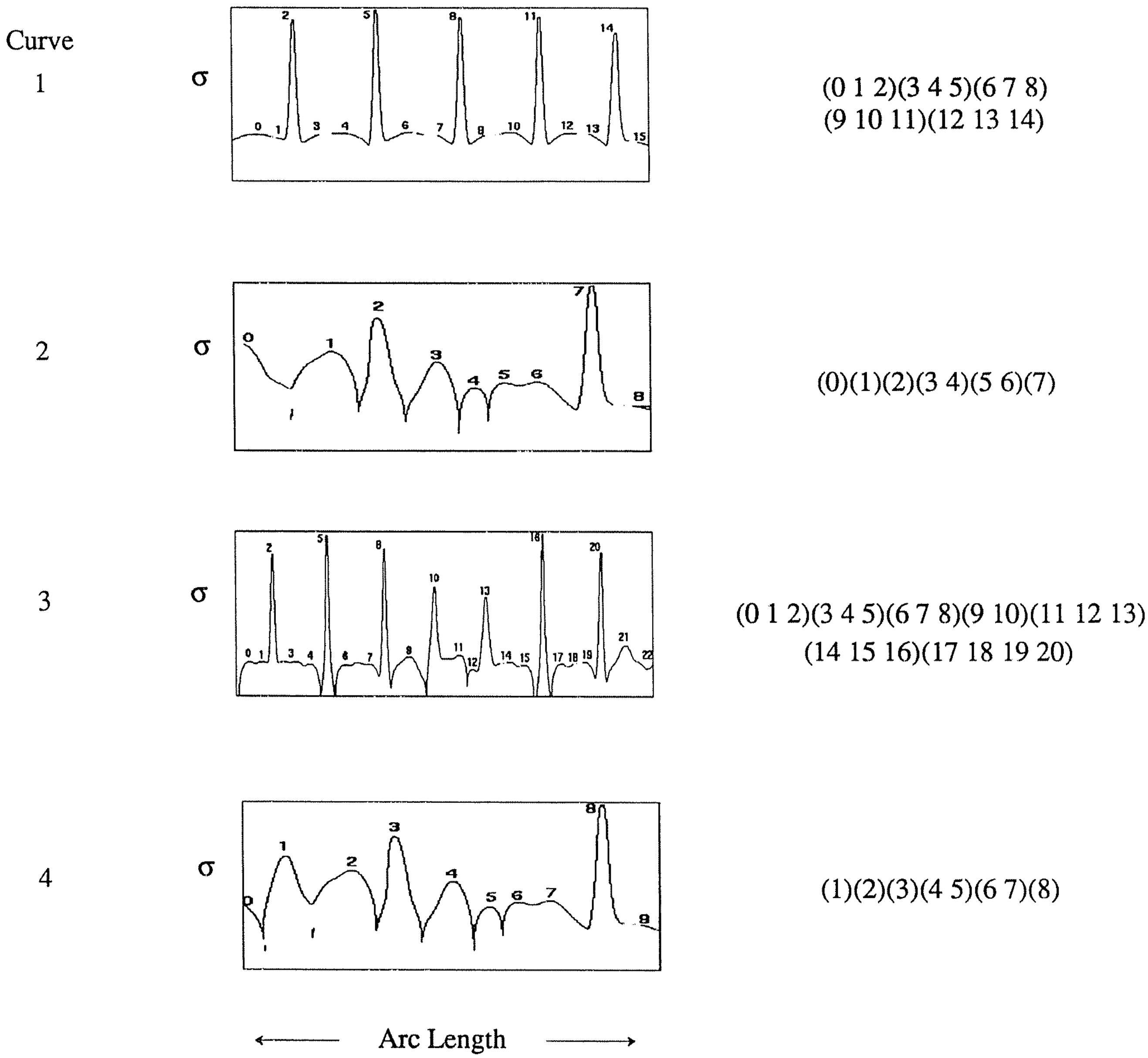


Figure 3.10. Curvature scale spaces images for the spatiotemporal curves of the flapping wings image sequence. The repeating patterns are shown on the right.

The cost between two features is computed by translating one feature so that its top corresponds to the top of the second feature. A measure of the difference of the two features is then computed. The cost is normalized so that large and small features are treated equally. But since it is desirable to match large features initially, the cost is reduced as a function of the size of the features being matched.

In the initial creation of the nodes, the cost is further modified so as to favor following paths that initially match temporally-adjacent periods rather than paths that match periods father apart.

4. Results

Figure 3.9 shows the curvature scale-space images for the spatiotemporal curves of the translating box image sequence. Next to each scale-space image are the detected repeating patterns. The same is shown for the flapping wings image sequence in Figure 3.10.

Table 4.1 shows the approximate frames when the cycles start. After the best cycles were detected, a threshold was used to eliminate weak matches, e.g. curves 1 and 3 for the flapping wings sequence.

Table 4.1

Results of Cycle Detection			
Translating Box Image Sequence		Flapping Wings Image Sequence	
Curve	Starting Frame Numbers	Curve	Starting Frame Numbers
1	8 17 29 36 49 60	1	5 11 18 24 30
2	10 20 30 39 49 59	2	No cycles found
3	9 17 27 36 47 59	3	3 7 12 16 20 24 29
4	9 19 30 39 49 60	4	No cycles found
5	10 21 33 41 51		
6	9 19 30 38 48 58		
7	9 19 30 37 48 58		

5. Concluding Remarks

In this paper we defined a procedure for detecting cyclic motion based on tracking curvature extrema in spatiotemporal images. A scale-space representation was then used to efficiently detect repeating patterns. The results show that it is possible to detect cyclic motion without first identifying the object(s) undergoing the motion. However, our results depended on two assumptions: objects do not rotate in depth and do not become occluded or disoccluded.

The first assumption was necessary because point tracking was used to detect ST-curves. This assumption can be dropped if we use principle curvature directions to detect ST-curves. Using the Implicit Function Theorem we can derive the First Normal Form of the spatiotemporal surface using only the coordinate points, i.e. we do not need a parameterized or analytic description of the spatiotemporal surface. It is also possible to fit a surface to the neighborhood of points known to be on the ST-surface [Besl86]. In both cases we can recover the coefficients of the first fundamental form which then allow us to calculate the principle curvature directions for any point on the surface. The ST-curves can then follow either of the principle curvature directions with the constraint that the ST-curve move forward in time. ST-curves defined in this manner are guaranteed to be consistent.

The assumption that moving objects do not become occluded or disoccluded was made because a coarse-level analysis of spatiotemporal volumes and surfaces is not yet developed. We have developed a motion representation hierarchy that consists of a low, intermediate and high level. Cyclic motion description is an example of an intermediate-level motion description. ST-surface and ST-volume interaction is another intermediate-level motion description. These interactions describe, among other things, occlusion and disocclusion between objects. Once this intermediate level is implemented, we can drop our assumption of no occlusion or disocclusion in the image sequence.

When the repeating cyclic motion is as simple as the examples shown in this paper, the curvature scale-space consists of spike-like features. Rather than using a uniform cost search, we could alternately compute the Fourier transform of the scale-space image and recover the frequency of the peaks. While this will work for simple repeating curvature values, it is not clear how well it will work for more complicated repeating motions.

References

[Akit84] Akita, K., ''Image sequence analysis of real world human motion,'' *Pattern Recognition*, **17**, pp. 73-83 (1984).

[Aloi88] Aloimonos, J., I. Weiss, and A. Bandyopadhyay, ''Active vision,'' *International Journal of Computer Vision*, pp. 333-356 (1988).

[Ande85] Anderson, C. H., P. J. Burt, and G. S. VanDerWal, ''Change detection and tracking using pyramid transform techniques,'' *Proc. SPIE Intelligent Robots and Computer Vision*, **579**, pp. 72 - 78 (1985).

[Bake88] Baker, H. H. and R. C. Bolles, ''Generalizing epipolar-plane image analysis of the spatiotemporal surface,'' *Proc. Computer Vision and Pattern Recognition*, pp. 2-9 (1988).

[Barr81] Barrow, H. G. and J. M. Tenenbaum, ''Computational vision,'' *Proceedings of the IEEE*, **69**, pp. 572-595 (May, 1981).

[Besl86] Besl, P. J. and R. C. Jain, ''Invariant surface characteristics for 3D object recognition in range images,'' *Computer Vision, Graphics, and Image Processing*, **33**, pp. 33-80 (1986).

[Boll87] Bolles, R. C., H. H. Baker, and D. H. Marimont, ''Epipolor-plane image analysis: An approach to determining structure from motion,'' *International Journal of Computer Vision*, **1**, pp. 7-55 (1987).

[Bolt94] Bolton, T. L., ''Rhythm,'' *American Journal of Psychology*, **6**, pp. 145-238 (1894).

[Coxe61] Coxeter, H. S. M., *Introduction to Geometry,* Wiley, New York (1961).

[DoCa76] DoCarmo, M., *Differential Geometry of Curves and Surfaces,* Prentice-Hall, Englewood Cliffs, New Jersey (1976).

[Frai75] Fraisse, P., ''Is rhythm gestalt?,'' pp. 227-232 in *Gestalttheorie in der modernen Psychologie*, ed. E. L. Kemmler and M. Stadler, (1975).

[Frai78] Fraisse, P., ''Time and rhythm perception,'' pp. 203-254 in *Handbood of Perception Volume VIII*, ed. E. C. Carterette and M. P. Friedman,Academic Press, New York (1978).

[Gabr73] Gabrielsson, A., ''Adjective ratings and dimension analysis of auditory rhythm patterns,'' *Scandinavian Journal of Psychology*, **14**, pp. 244-260 (1973).

[Godd88a] Goddard, N. H., ''Recognizing animal motion,'' *Proceedings of Image Understanding Workshop*, pp. 938-944 (1988).

[Godd88] Goddard, N. H., ''Representing and recognizing event sequences,'' *Proc. CMU Summer Connectionist Workshop*, (1988).

[Godd89] Goddard, N. H., ''The interpretation of visual motion: Recognizing moving light displays,'' *Proc. IEEE Workshop on Motion*, (March, 1989).

[Goul89] Gould, K. and M. Shah, ''The trajectory primal sketch: A multi-scale scheme for representing motion characteristics,'' *Proc. Computer Vision and Pattern Recognition*, pp. 79-85 (1989).

[Hogg83] Hogg, D., ''Model-based vision: A program to see a walking person,'' *Image and Computer Vision Computing*, **1**, pp. 5-20 (1983).

[Jain79] Jain, R. and H. H. Nagel, ''On the analysis of accumulative difference pictures from image sequences of real world scenes,'' *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **1**, pp. 206-214 (1979).

[Jain79a] Jain, R., W. N. Martin, and J. K. Aggarwal, ''Segmentation through the detection of changes due to motion,'' *Computer Vision, Graphics and Image Processing*, **11**, pp. 13-34 (1979).

[Jain81] Jain, R., ''Dynamic scene analysis using pixel-based processes,'' *IEEE Computer*, **14**, pp. 12-19 (1981).

[Jain83] Jain, R., ''Extraction of motion information from peripheral processes,'' *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **3**, pp. 489-503 (1983).

[Jain88] Jain, R., ''Dynamic vision,'' *Proc. 9th International Conference on Pattern Recognition*, pp. 226-235 (1988).

[Jaya83] Jayaramamurthy, S. N. and R. Jain, ''An approach to the segmentation of textured dynamic scenes,'' *Computer Vision, Graphics and Image Processing*, **21**, pp. 239-261 (1983).

[Joha73] Johansson, G., ''Visual perception of biological motion and a model for its analysis,'' *Perception and Psychophysics*, **14**, pp. 201-211 (1973).

[Koff09] Koffka, K., ''Experimental-untersuchunger zur lehre vom rhythmus,'' *Zeitschrift fur Psychologie*, **52**, pp. 1-109 (1909).

[Lee88] Lee, J. S. and C. Lin, ''A novel approach to real-time motion detection,'' *Proc. Computer Vision and Pattern Recognition*, pp. 730 - 735 (1988).

[Leun87] Leung, M. K. and Y. H. Yang, ''Human body motion segmentation in a complex scene,'' *Pattern Recognition*, **20**, pp. 55-64 (1987).

[Leun87a] Leung, M. K. and Y. H. Yang, ''A region based approach for human body motion analysis,'' *Pattern Recognition*, **20**, pp. 321-339 (1987).

[Leyt88] Leyton, M., ''A process-grammar for shape,'' *Artificial Intelligence*, **34**, pp. 213 - 247 (March 1988).

[Liou89] Liou, S. P. and R. C. Jain, ''Motion detection in spatio-temporal space,'' *Computer Vision, Graphics, and Image Processing*, **45**, pp. 227-250 (1989).

[Mokh86] Mokhtarian, F. and A. Mackworth, ''Scale-based description and recognition of planar curves and two-dimensional shapes,'' *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **8**, pp. 34-43 (January, 1986).

[Mokh88] Mokhtarian, F., ''Multi-scale description of space curves and three-dimensional objects,'' *Proc. Computer Vision and Pattern Recognition*, pp. 298-303 (1988).

[Rubi85] Rubin, J. and W.A. Richards, ''Boundaries of visual motion,'' *AI Memo 835*, MIT, (April 1, 1985).

[Witk87] Witkin, A., ''Scale-space methods,'' pp. 973-980 in *Encyclopedia of Artificial Intelligence*, ed. S. Shapiro, Wiley, New York (1987).

[Yuil83] Yuille, A. L. and T. Poggio, ''Fingerprint theorems for zero crossings,'' *AI Memo 730*, MIT, (1983).

Appendix 1: Proof Sketch of Theorem 3.1

Theorem 3.1: If a solid object in the scene displays cyclic motion such that the cyclic motion is preserved under projection, then the spatiotemporal surfaces that correspond to intervals of cyclic motion will be piecewise isomorphic.

Proof Sketch: Consider a spatiotemporal surface such that there exist ridges on the surface and every point on a ridge has the same value for the time coordinate. We call this ridge a *temporal ridge* since it results from a nonsmooth change of velocity of an object at some time. Define a *patch* of an ST-surface as the surface between two temporal ridges. Two surfaces are piecewise isomorphic if the first patch of surface 1 is isomorphic to the first patch of surface 2, the second patches are isomorphic, etc. Let C be a closed curve in the image that is the result of the projection into the x-y plane of a solid object undergoing cyclic motion. Let

$$x = f(u), \qquad y = g(u), \qquad a < u < b, \qquad f(u) > 0, \qquad g(u) > 0$$

be a parameterization for C. As the object in the scene moves the curve in the image also moves. We will now parameterize the surface, S, generated by C moving in the image. The resulting surface will exist in x-y-t space. We will break the types of movement of C into three cases.

Case I: *C* Translates

Let v_x and v_y be the translational components of the curve in the x and y directions, respectively. Let d_x and d_y be the displacement of C in the x and y directions, respectively. We obtain the following map

$$\mathbf{H}(u, t) = (f(u) + tv_x + d_x, g(u) + tv_y + d_y, t)$$

from the open set $U = \{(u, t) \in \mathbb{R}^2; a < u < b; c < t < d; c, d \geq 0\}$ into S.

Case II: *C* Rotates

Let (i, j) be the point about which C rotates. Let ω be the angular velocity of C. Let θ be the angle that a point on C has rotated through. We obtain the following map

$$H(u, t) = (|i - f(u)|\cos(t\omega), |j - g(u)|\sin(t\omega), t)$$

from the open set $U = \{(u, t) \in \mathbb{R}^2;\ a < u < b;\ c < t < d;\ c, d \geq 0\}$ into S. If necessary, displacement terms can be added to the x and y components.

Case III: *C* Translates and Rotates

We simply combine the components of the parameterization for Cases II and III.

It is possible that over some time interval, C will undergo a combination of these movements, changing parameters with time. For example, C may translate at one speed for a time, continue translating at a different speed, then rotate about some point for a time, and finally rotate about a different point for a time. But we can simply break this sequence of movements into separate parameterizations for each translation, rotation, non-movement and combination of translation and rotation. One parameterization starts up where the previous one leaves off to give a continuous surface. Each of these parameterizations define a piece of the surface. It is these pieces that will be isomorphic. In the case where C accelerates, another parameterization can be considered or the ST-surface generated by C while accelerating can be treated as a series of translation and rotation changes.

Two surfaces are piecewise isomorphic if there exists a bijective function mapping one patch to another patch. We need to show that given two surface patches, one patch generated by C during one cycle of motion, the other patch generated by C during a corresponding cycle of motion, the two patches are isomorphic.

We assumed that the cyclic motion of a rigid object is preserved under projection. This means that for any point on the object, it displays cyclic motion under projection. Let S_1 and S_2 be two surfaces patches that were generated by C at corresponding times in the cyclic motion of C. Let H_{S_1} and H_{S_2} be the parameterizations of S_1 and S_2, respectively. Let patch S_i start at t_i. Since there are three different types of parameterizations we need to show that for any of these parameterizations, corresponding patches are isomorphic.

Case I: Translation

Given the parameterization for two translating patches:

$$\mathbf{H}_{\mathbf{S_1}}(u,t) = (f(u) + tv_x + d_{x_1}, g(u) + tv_y + d_{y_1}, t)$$
$$\mathbf{H}_{\mathbf{S_2}}(u,t) = (f(u) + tv_x + d_{x_2}, g(u) + tv_y + d_{y_2}, t)$$

we need to show that there exists a bijective mapping from S_1 to S_2. The function:

$$h([x,y,t]) = [x,y,t] + [d_{x_2} - d_{x_1}, d_{y_2} - d_{y_1}, t_2 - t_1]$$

where t_i is the start time of patch S_i, is a bijection that maps points from S_1 to S_2. So S_1 and S_2 are isomorphic.

Case II: Rotation

The definition of cyclic motion uses the curvature of the path of motion rather than the position of the path. Suppose an object is undergoing cyclic motion such that this cyclic motion is preserved under projection. All this says is that the curvature is repeating. Let t_1 and t_2 be two corresponding times in different periods. Since these times correspond we know that the curvature at corresponding points is the same. Given that the projection is rotating in the image plane, we need to show that there is only one possible center of rotation for C that could give rise to the curvature values of the points of C over time.

The curvature of a point is equal to the inverse of the distance from that point to the the center of rotation. Given that the curvature of point p_i on C is k_i, its distance from the center of rotation is $1/k_i$. We need to show that there is only one center of rotation point such that the distance from that center of rotation point to p_i is $1/k_i$ for all p_i on C. We will show this by assuming that two centers of rotation exist and show that C must then be a straight line. Let (x_1, y_1) and (x_2, y_2) be the two centers of rotation. Position the coordinate system so that y_1 and y_2 are both zero and the y-axis is a perpendicular bisector of the line segment joining (x_1, y_1) and (x_2, y_2). See Figure A1.1.

Let (x_p, y_p) be a point distance d away from (x_1, y_1) and (x_2, y_2). We need to show that (x_p, y_p) lies on the y-axis. In other words, we have to show that $x_p = 0$. The distances from (x_1, y_1) to (x_p, y_p) and from (x_2, y_2) to (x_p, y_p) are:

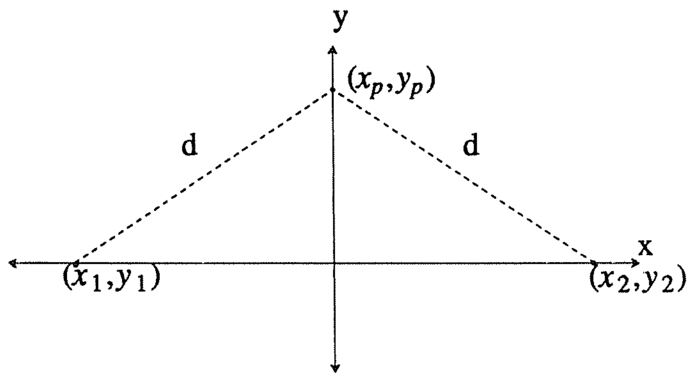


Figure A1.1. The distance from (x_1,y_1) to (x_p,y_p) equals the distance from (x_2,y_2) to (x_p,y_p).

$$d = \sqrt{(x_p - x_1)^2 + (y_p - y_1)^2} \qquad d = \sqrt{(x_p - x_2)^2 + (y_p - y_2)^2}$$

respectively. Setting these two equal and solving for x_p we find that x_p equals 0. So we have shown that if there are two possible center of rotation points then C must be a straight line. Since C is assumed to be a closed curve, i.e. not a straight line, there can only be one center of rotation.

Let (i, j) be the center of rotation point. There are two directions that C can rotate around (i, j). But the curvature of the paths of the points of C for rotating one direction will be the negation of the curvature values if C rotates in the other direction. So there is actually only one direction that will give the correct curvature values for the points of C. We have now shown that C must rotate in a unique direction around a unique point. Given the parameterization for two rotating patches S_1 and S_2:

$$\mathbf{H}_{S_1}(u,t) = (|i - f(u)|\cos(t\omega) + |j - g(u)|\sin(t\omega),\ -|i - f(u)|\sin(t\omega) + |j - g(u)|\cos(t\omega),\ t)$$

$$\mathbf{H}_{S_2}(u,t) = (|i - f(u)|\cos(t\omega) + |j - g(u)|\sin(t\omega),\ -|i - f(u)|\sin(t\omega) + |j - g(u)|\cos(t\omega),\ t)$$

we need to show that there exists a bijective function from S_1 to S_2. The function:

$$h([x,y,t]) = [x,y,t] \begin{bmatrix} \cos((t_1 - t_2)\omega) & -\sin((t_1 - t_2)\omega) & 0 \\ \sin((t_1 - t_2)\omega) & \cos((t_1 - t_2)\omega) & 0 \\ 0 & 0 & t_1 - t_2 \end{bmatrix}$$

where t_i is the start time of patch S_i, is a bijection that maps points from S_1 to S_2. So S_1 and S_2 are isomorphic.

Case III: Translation and Rotation

We combine the translation and rotation of Case II and Case III to show that S_1 is isomorphic to S_2. Since C is translating, the center of rotation of C is translating also. But it remains unique for all t.

This shows that S_1 is isomorphic to S_2 in all cases. So the ST-surfaces generated by the cyclic motion of C are piecewise isomorphic.

□