

# Affine Surface Reconstruction By Purposive Viewpoint Control

Kiriakos N. Kutulakos

kyros@cs.rochester.edu

Department of Computer Sciences  
University of Rochester  
Rochester, NY 14627-0226 USA

## Abstract

We present an approach for building an affine representation of an unknown curved object viewed under orthographic projection from images of its occluding contour. It is based on the observation that the projection of a point on a curved, featureless surface can be computed along a special viewing direction that does not belong to the point's tangent plane. We show that by circumnavigating the object on the tangent plane of selected surface points, we can (1) compute two orthogonal projections of every point projecting to the occluding contour during this motion, and (2) compute the affine coordinates of these points. Our approach demonstrates that affine shape of curved objects can be computed directly, *i.e.*, without Euclidean calibration or image velocity and acceleration measurements.

## 1 Introduction

There has been a growing interest recently in performing tasks such as recognition [1], navigation [2], and shape recovery [3–6] using one or more uncalibrated cameras. One feature of these approaches is that they avoid sensitivity to errors in camera position and motion estimates by recovering non-Euclidean scene representations for which such measurements are unnecessary. Unfortunately, while techniques have been developed for recovering non-Euclidean representations of 3D point sets whose projection can be tracked across frames [1, 3, 7] (e.g., corners, surface markings, and discontinuities), the recovery of such representations for unknown curved objects with sparsely-distributed markings and discontinuities remains open.

In this paper we employ a mobile observer that *purposefully* controls its motion in order to build an affine representation of a curved object from images of its occluding contour. The approach is motivated by the *purposive* [8] and *animate* [9] vision paradigms, and combines elements from the study of the occluding contour of smooth surfaces

[10] and previous work on affine reconstruction [1, 4, 7]. We show that by appropriately controlling the observer's motion during a circumnavigation of the object, we can compute the affine coordinates of every surface point projecting to the occluding contour during this motion. The assumptions are that (1) the object is viewed under orthographic projection, (2) the observer's motion is planar, (3) a correspondence can be established between occluding contour curves in consecutive images, and (4) the projection of at least four points on the object's surface can be tracked across frames. These assumptions enable an analysis similar to the one of Weinshall and Tomasi [4] and allow affine shape to be recovered from the occluding contour without the need for Euclidean calibration or image velocity and acceleration measurements. Furthermore, we show that by selecting the observer's motion plane to be tangent to an automatically-selected point on the surface, the fourth assumption can be reduced to simply identifying four occluding contour points with parallel tangents in the initial image and tracking those points across frames.

The occluding contour of a smooth surface is defined as the projection of the visible rim, which is the one-dimensional set of visible surface points at which the line of sight is tangent. This set depends on the shape of the surface and the observer's viewpoint. Under continuous observer motion the visible rim "slides" over the surface, affecting the geometry of the occluding contour [10–12]. It is well-known that the deformation of the occluding contour uniquely determines surface shape. At issue is how to recover shape accurately. For example, previous methods measure the speeds and accelerations of image curves and relate them to the speeds and accelerations of the observer [11, 13–15]. As a result, sensitivity to camera calibration and motion measurement errors remains an important issue, especially in active vision systems where one or more of the camera's intrinsic and extrinsic parameters may change during the reconstruction process [16].

While the visible rim's dependency on viewpoint has proved very useful for extracting Euclidean representations of curved objects, this dependency is the main reason existing techniques for recovering the affine or projective shape

---

The support of the National Science Foundation under Grant No. CDA-9503996, of the University of Maryland under Subcontract No. Z840902, and of Honeywell under Research Contract No. 304931455 are gratefully acknowledged.

of such objects are not applicable. A basic requirement for extracting non-Euclidean representations of an arbitrary 3D point set is that the projections of all points in the set are known at at least two viewpoints. Unfortunately, the sliding motion of the visible rim implies that even when correspondence between occluding contour curves can be established in images taken at different viewpoints, their points are not, in general, projections of the same points on the surface. Previous approaches have thus focused on extracting invariant descriptions of curved objects from just a single image of their occluding contour. These approaches are applicable only to restricted classes of objects (e.g., generalized cylinders [17], or algebraic surfaces [18]), where the sparse information provided by the contour is sufficient to uniquely determine the object’s affine or projective shape.

In this paper we show that when viewpoint is controllable, the affine coordinates of every point projecting to the visible rim can be computed *directly*, i.e., without computing the visible rim’s motion over the surface or the observer’s motion parameters. We achieve this by controlling viewpoint on a single plane so that the projection of every surface point over which the visible rim slides can be determined along two viewing directions on that plane. Once these projections are determined, the points’ affine coordinates are computed using existing reconstruction methods [4, 7].

The main question we address is how to compute the projection of a point on an unknown curved object along a direction where the point does not project to the occluding contour. To achieve this we use the following observation: Suppose there is a special point on the observer’s motion plane whose projection can be computed along every direction on that plane. Then, for every surface point on that plane there is a unique viewing direction at which the point’s projection coincides with the projection of the special point. So if we can somehow adjust our viewing direction to align it with that direction, we can compute the point’s projection even though the point may not project to the occluding contour. In this paper we show that such a viewing direction adjustment is indeed possible. We develop a deterministic strategy for performing this viewing direction adjustment for every surface point that can project to the occluding contour. The strategy is driven directly by the image data and relies on maximizing or minimizing a specific geometric quantity (distance of the special point’s projection from the contour). Furthermore, we show how the affine coordinates of the special points employed by the strategy can be computed from the object’s occluding contour.

The significance of our method lies in the use of purposeful observer motion to maintain specific geometric relationships with the viewed object in order to derive an affine representation of the object’s surface. Consequently, observer motion is not used to merely change the shape of

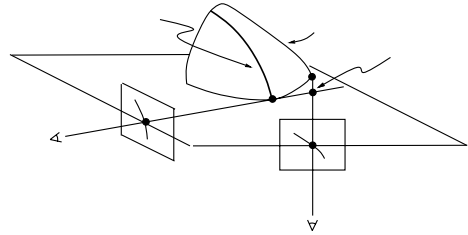


Figure 1: Point correspondences induced by the epipolar geometry. Point  $q(t_2)$  is the point corresponding to  $q(t_1)$ . When  $p(t_1)$  is on the visible rim and its tangent plane is not an epipolar plane,  $q(t_2)$  is the projection of  $p(t_2)$ , not  $p(t_1)$ . The lines through  $p(t_1), p(t_2)$  along  $\xi(t_1), \xi(t_2)$ , respectively, intersect at a *virtual point*  $v$  (Section 5).

the occluding contour (as in previous approaches), but it is used to change it in a well-defined way, factoring out the need for differential image measurements and for Euclidean calibration.

## 2 Viewing Geometry

Let  $S$  be a smooth, oriented surface in  $\mathbb{R}^3$ , viewed under orthographic projection along a viewing direction<sup>1</sup>  $\xi$ . The *visible rim* of  $S$  is the set of visible points  $p$  whose tangent plane,  $T_p(S)$ , contains a line parallel to  $\xi$ . The *occluding contour* of  $S$  is the projection of the visible rim on the image plane. When viewpoint changes along a continuous path, the visible rim curves slide over the surface. Our goal is to obtain a representation for all points on the surface over which the visible rim slides. This operation requires a correspondence between points on the surface at different viewpoints, and between points in images. We use the epipolar plane correspondences [11, 13, 14] for this purpose.

In particular, suppose the viewing direction changes according to  $\xi(t)$  on a motion plane with normal  $N_E$ . This motion defines a family of planes parallel to the motion plane called the *epipolar planes*. If  $B(t_1)$  is an occluding contour curve at time  $t_1$ , the epipolar plane correspondence matches a point  $q(t_1) \in B(t_1)$  to the intersection of  $B(t_2)$  with the epipolar plane through  $q(t_1)$  (Figure 1). This correspondence induces a correspondence between points belonging to the visible rim curve that projects to  $B(t)$ . In the following, if  $p$  is a surface point projecting to  $q$  at the initial viewpoint, we denote by  $p(t)$  and  $q(t)$  the points corresponding to  $p$  and  $q$ , respectively, at viewpoint  $\xi(t)$ .

When viewpoint moves on a single plane, the point  $p(t)$

<sup>1</sup>In the following we use the terms “viewing direction” and “viewpoint” interchangeably.

slides over a curve in the intersection of the surface with the epipolar plane through  $p(0)$ . In general, this intersection may contain points that do not lie on the visible rim for any viewpoint on the motion plane. Given an epipolar plane  $E$ , the *potential* visible rim points on  $E$  are the points of  $S \cap E$  that lie on the visible rim for some viewpoint in the motion plane. Our goal is to control viewpoint on a single plane in order to build a representation for all potential visible rim points.

## 2.1 Affine Point Representations

Previous work on recovering surface shape from the occluding contour considered how the surface can be represented as a collection of local surface patches that are combined in a Euclidean frame of reference [11, 14]. Unlike previous approaches, we derive a surface representation consisting of a collection of points that are put in an object-centered, *affine* reference frame. The basic principles behind this representation are briefly reviewed next.

Let  $p_1, \dots, p_n \in \mathbb{R}^3, n \geq 4$ , be a collection of points at least four of which are not coplanar. An *affine representation* of those points is a representation that does not change if the same non-singular linear transformation (e.g., translation, rotation, scaling) is applied to all the points. Affine representations consist of three components: The *origin*, which is one of the points  $p_1, \dots, p_n$ ; the *affine basis points*, which are three points from the collection that are not coplanar with the origin; and the *affine coordinates* of the points  $p_1, \dots, p_n$ .

The key advantage of affine point representations is that they can be recovered from two images of the points without any information about the camera’s calibration or about its motion between the two images. Specifically, we use the following two properties of affine point representations:<sup>2</sup>

**Property 1 (Re-Projection Property)** When the projection of the origin and basis points is known along a viewing direction  $\xi(t_m)$ , we can compute (1) the epipolar correspondences, and (2) the projection of a point  $p$  from its affine coordinates using

$$\begin{bmatrix} u_p^m \\ v_p^m \end{bmatrix} = \begin{bmatrix} u_{b_1}^m & u_{b_2}^m & u_{b_3}^m \\ v_{b_1}^m & v_{b_2}^m & v_{b_3}^m \end{bmatrix} \mathbf{A}_p + \begin{bmatrix} u_{p_o}^m \\ v_{p_o}^m \end{bmatrix} \quad (1)$$

where  $(u_p^m, v_p^m)$  is the projection of  $p$ ,  $(u_{p_o}^m, v_{p_o}^m)$  is the projection of the origin, and  $\mathbf{A}_p$  is the vector collecting  $p$ ’s affine coordinates.

**Property 2 (Affine Reconstruction Property)** The affine coordinates of  $p_1, \dots, p_n$  can be computed using Eq. (1) when their projection along at least two viewing directions is known.

<sup>2</sup>See [1, 4, 7, 19] for the related theory, details on these properties, and for algorithms to compute the affine representation of a collection of points.

## 3 Affine Shape From Contour

The Affine Reconstruction Property allows us to recover the affine coordinates of a set of 3D points from two images of the points. Unfortunately, this property cannot be used directly for computing an affine representation of a curved and featureless surface from its occluding contour. The reason is that corresponding occluding contour points in two images are not projections of the same point on the surface (Figure 1). The main idea of our approach is that for any potential visible rim point there is always a pair of orthogonal viewing directions along which the point’s projection can be determined. Hence, we use an active observer that purposefully changes viewpoint in order to reach these viewing directions. Once these viewing directions are reached and the point’s projection is determined, its affine coordinates can be computed from Eq. (1). In the following we assume that the projection of four basis points has been identified *a priori* and can be tracked across frames. In addition to providing an affine basis frame, this assumption allows epipolar correspondences to be established. This assumption is lifted in Section 4.

Suppose the observer is constrained to move on a single plane. Let  $p$  be a potential visible rim point, and let  $E$  be the epipolar plane through  $p$ . Point  $p$  projects to the occluding contour only for viewing directions in  $T_p(S) \cap E$ ; if  $T_p(S)$  and  $E$  do not coincide (which is the case for all but a collection of isolated points), this intersection is a line  $l_p$ . In order to compute  $p$ ’s affine coordinates using Eq. (1) we determine  $p$ ’s projection along  $l_p$  and along an additional viewing direction that is perpendicular to  $l_p$ . We use the following simple observation which suggests an *indirect* method for determining the projection of  $p$  along such a viewing direction (Figure 2(a)):

**Observation 1** Let  $l_p^\perp$  be the line through  $p$  that is perpendicular to  $l_p$  and lies on  $p$ ’s epipolar plane. If  $v$  is a point on  $l_p^\perp$ , the projections of  $p$  and  $v$  coincide for viewing directions parallel to  $l_p^\perp$ .

Observation 1 tells us that  $p$ ’s projection along line  $l_p^\perp$  can be determined even though  $p$  does not project to the occluding contour from such a viewpoint: It suffices to know the projection of any point along that line. In the following, we refer to viewing directions parallel to  $l_p$  and  $l_p^\perp$  as  $p$ ’s *tangential* and *normal* directions, respectively, and to  $p$ ’s projection along these directions as its *tangential* and *normal* projections.

Observation 1 imposes two important requirements for indirectly determining the normal projection of a potential visible rim point  $p$ : The observer must be capable of (1) defining a point  $v$  that belongs to  $l_p^\perp$ , and (2) computing  $v$ ’s projection along the normal direction of  $p$ . The key idea

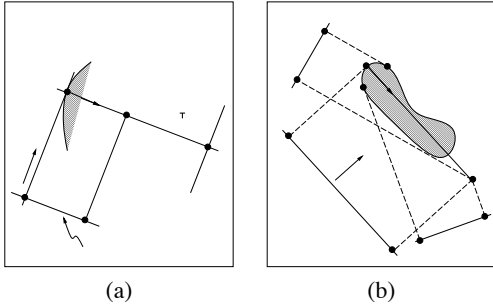


Figure 2: (a) Computing the projection of a potential visible rim point  $p$  along two orthogonal viewing directions. Point  $p$  projects to the occluding contour along  $l_p$ .  $\mathbf{n}(p)$  is the normal of  $S \cap E$  at  $p$ . Even though  $p$  does not project to the occluding contour along viewing directions that are not parallel to  $l_p$ , its projection along  $l_p^\top$  can be determined indirectly from the projection of point  $v$ . (b) Reaching the tangential direction of a point  $p \in P_v$ .  $p$  belongs to the visible rim along  $\xi(t^*)$ , where the distance between  $q(t)$  and  $v$ 's projection are maximized.

used in our approach is that if we simply want to compute the tangential and normal projection of an unspecified point in  $S \cap E$ ,  $v$  can be almost arbitrary. In particular, suppose that the observer circumnavigates the object either clockwise or counter-clockwise by changing viewing direction on a single plane according to  $\xi(t)$ ,  $0 \leq t \leq T$ . We use the following theorem<sup>3</sup> (Figure 2(b)):

**Theorem 1** Let  $v$  be an arbitrary point on  $E$  and let  $r(t)$  be its projection along  $\xi(t)$ . Suppose  $P_v$  is the set of potential visible rim points  $p$  for which  $v$  is along  $l_p^\top$ . (1)  $P_v$  is equal to the non-empty set of potential visible rim points whose distance from  $v$  is locally extremal. (2) Suppose  $q(t)$  is an occluding contour point on  $E$  when the viewing direction is  $\xi(t)$ . The image separation  $\|q(t) - r(t)\|$  exhibits a local extremum at  $t = t^*$  if and only if the visible rim point projecting to  $q(t^*)$  is in  $P_v$ .

So assume that we can select a point  $v$  on the epipolar plane  $E$  and can compute its projection along arbitrary viewing directions on  $E$ . The first part of Theorem 1 guarantees the existence of at least one potential visible rim point whose normal projection can be determined indirectly from the projection of  $v$ . Furthermore, the theorem's second part gives us a way to reach the viewpoint where a point  $p \in P_v$  belongs to the visible rim, and to determine  $p$ 's tangential projection: It suffices to circumnavigate the object on the motion plane until the distance between an occluding contour point  $q(t)$  on  $E$  and  $v$ 's projection is either minimized or maximized.  $p$ 's projection will be the point  $q(t)$ . Once

$p$ 's tangential direction is reached, its normal direction can be reached by performing a viewing direction change of  $\frac{\pi}{2}$  radians;  $p$ 's normal projection can then be determined indirectly from the projection of  $v$ .

The potential visible rim points whose normal projections can be computed in this fashion are not under the observer's control; they depend on  $v$  as well as the shape of the surface. The following corollary to Theorem 1 goes a step further, showing that if we can compute the projection of every point on a curve surrounding the surface at every viewpoint, we can compute the tangential and normal projection of every potential visible rim point on the observer's motion plane:

**Corollary 1** Let  $\alpha$  be a closed curve on the epipolar plane  $E$ . If  $\alpha$  contains  $S \cap E$  in its interior and  $p$  is a potential visible rim point on  $E$ , there is at least one point  $v$  on  $\alpha$  such that  $p \in P_v$ .

## 4 Active Affine Shape Recovery

Our goal is to control the observer's viewpoint in order to recover the affine coordinates of all potential visible rim points. We do this by (1) selecting a motion plane, (2) selecting an affine basis, and (3) circumnavigating the object on that plane so that the tangential and normal projections of all potential visible rim points can be determined. The motion plane and affine basis selection problems were studied in [12]: If four occluding contour points with parallel tangents can be identified in the image and the observer's motion plane is selected parallel to the points' tangents, those points define a valid affine basis.

The basic step of our method is to determine the tangential and normal projections of all potential visible rim points on a single epipolar plane; this step can then be applied in parallel to all epipolar planes intersecting the initial image. Below, we first outline the main ideas of this step, and in Sections 5 and 6 present more details.

Suppose that the observer's motion plane has been selected. Let  $q$  be a point on the occluding contour at the initial image, and let  $E$  be the epipolar plane through  $q$ . Observation 1 tells us that if we can define a point  $v$  on  $E$  and can compute its projection along arbitrary viewing directions on the selected motion plane, we can compute the tangential and normal projection of a potential visible rim point  $p \in P_v$  by moving to its tangential and normal directions. Furthermore, Corollary 1 says that if we repeatedly perform this process by selecting the points  $v$  from an *a priori* defined curve on  $E$  surrounding the object, we will obtain increasingly more complete representations of the potential visible rim points on  $E$ . We consider how such a representation can be built by circumnavigating the object.

Viewing directions on  $E$  can be thought of as positions on a unit circle  $C$ , defined by the intersection of the unit sphere with  $E$ . As the observer changes viewing direction

<sup>3</sup>See [20] for a proof.

on  $E$  the corresponding point moves on  $C$ . Let  $\alpha$  be an *a priori* defined curve on  $E$  surrounding the object, and let  $v_1, \dots, v_n \in \alpha$ . Since  $C$  represents the set of all possible viewing directions on  $E$ , the tangential direction of every potential visible rim point in  $P_{v_i}, i = 1, \dots, n$ , corresponds to a position on  $C$ . Hence, to determine the tangential projection of some point  $p$  in a specific set  $P_{v_i}$ , it suffices to detect when the corresponding position on  $C$  is reached during a circumnavigation of the object. Theorem 1 allows us to perform this detection using a simple extremum-based method: The observer’s viewing direction will be aligned with the tangential direction of  $p$  when the distance between  $v_i$ ’s projection and an occluding contour point on  $E$  is either minimized or maximized.

Once  $p$ ’s tangential direction has been detected, its normal direction is determined by detecting when the observer’s viewing direction becomes perpendicular to  $p$ ’s tangential direction. This allows the tangential and normal directions of all potential visible rim points in  $P_{v_i}, i = 1, \dots, n$ , to be detected after total viewing direction change of at most  $\frac{5}{2}\pi$  radians. These considerations lead to the following strategy for computing the affine coordinates of potential visible points on a single epipolar plane:

### Active Affine Shape Recovery Strategy

- Step 1:** Select a motion plane  $E$  and an affine basis.
- Step 2:** Compute a continuous curve  $\alpha$  containing  $S \cap E$  in its interior, and sample it to obtain points  $v_1, \dots, v_n$ .
- Step 3:** Perform a small clockwise viewing direction change.
- Step 4:** Let  $q_1(t), \dots, q_m(t)$  be the occluding contour points on  $E$  along  $\xi(t)$ .
- Step 5:** For every  $v_i$  and every  $q_j(t)$ ,
  - a. compute the projection,  $r_i(t)$ , of  $v_i$  along  $\xi(t)$ ,
  - b. compute the distance  $\delta_{ij}(t) = \|r_i(t) - q_j(t)\|$ ,
  - c. if  $\delta_{ij}(t)$  is a local extremum of  $\delta_{ij}$ , the tangential direction of a point in  $P_{v_i}$  has been reached. Store the tuple  $\langle \mathbf{i}, \mathbf{q}_j(t), \xi(t) \rangle$  along with the projections of the affine basis points.
- Step 6:** If  $\xi(t)$  is perpendicular to the direction  $\xi(t')$  in a stored tuple  $\langle \mathbf{k}, \mathbf{q}_l(t'), \xi(t') \rangle$ , the normal direction of the point  $p \in P_{v_k}$  projecting to  $q_l(t')$  has been reached. Solve Eq. (1) with  $[u_p^1 \ v_p^1]^T = q_l(t')$  and  $[u_p^2 \ v_p^2]^T = r_k(t)$  to compute  $p$ ’s affine coordinates.
- Step 7:** Repeat Steps 3-6 until the total viewing direction change is at least  $\frac{5}{2}\pi$  radians.

In the next section we introduce the notion of *virtual points*, and briefly discuss how virtual points enable us to perform Step 6 of the Active Affine Shape Recovery Strategy. Section 6 then studies the extremum detection process of Step 5c in the presence of noise in image measurements.

## 5 Virtual Points

The Active Affine Shape Recovery Strategy described in the previous section made an important assumption: Given an epipolar plane  $E$ , the observer must be able to define points on  $E$  whose projection can be computed along viewing directions on  $E$ . We call such points *virtual*, because they are not required to correspond to any surface markings, or required to lie on the object’s surface (Figure 1):

**Definition 1 (Virtual Points)** If  $q(t_1), q(t_2)$  are corresponding occluding contour points and  $\xi(t_1)$  and  $\xi(t_2)$  are not parallel, there is a unique point  $v \in E$ , called a *virtual point*, whose projection along  $\xi(t_i)$  is  $q(t_i), i = 1, 2$ , and whose affine coordinates are computed by solving Eq. (1) with  $[u_p^i \ v_p^i]^T = q(t_i), i = 1, 2$ . Furthermore,  $v$  is the point of intersection of the ray through  $q(t_1)$  along  $\xi(t_1)$ , and the ray through  $q(t_2)$  along  $\xi(t_2)$ .

The affine coordinates of a virtual point on a given epipolar plane can be computed by selecting an occluding contour point on that plane, changing viewing direction, and applying the above definition.

In our approach, we use virtual points to detect orthogonal viewing direction adjustments without requiring viewer motion measurements (Step 6 of the Active Affine Shape Recovery Strategy). Specifically, suppose the observer is circumnavigating the object according to  $\xi(t)$  and consider the problem of detecting when  $\xi(t)$  becomes perpendicular to  $\xi(0)$ . We use the following observation which allows us to perform this detection using a simple maximization approach (Figure 3):

**Observation 2** Let  $v_1, v_2$  be two distinct virtual points along  $\xi(0)$ . The separation between the projections of  $v_1, v_2$  at  $\xi(t)$ , as a function of  $t$ , has exactly two maxima corresponding to viewing direction changes of  $\frac{\pi}{2}$  and  $\frac{3\pi}{2}$  radians relative to  $\xi(0)$ .

Observation 2 allows detection of orthogonal viewing direction changes by first computing the affine coordinates of two virtual points on a line parallel to the initial viewing direction and then detecting when their distance is maximized in the image (Figure 3).

Virtual points can also be used to define a planar curve enclosing the object without having to explicitly extract information about the object’s shape (Step 2 of the Active Affine Shape Recovery Strategy). We show in [20] that such a curve can be defined as the convex hull of a collection of virtual points whose coordinates are computed during a preliminary circumnavigation of the object.

## 6 Measurement Errors

Our Active Affine Shape Recovery Strategy relies on two extremum-finding operations: (1) Determining at which

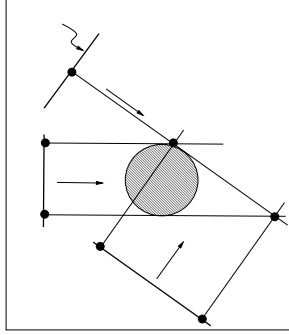


Figure 3: Detecting orthogonal viewing direction changes. The virtual points  $v_1, v_2$  are the intersections of the ray through  $q(0)$  along  $\xi(0)$  with the ray through  $q_1(\delta t)$  along  $\xi(\delta t)$ , and the ray through  $q_2(\delta t)$  along  $\xi(\delta t)$ , respectively. The orthogonal direction  $\xi(t^*)$  is reached when the separation of the virtual points' projection is maximized.

viewpoint during an object's circumnavigation the projection of a virtual point is at an extremal distance from the occluding contour (Step 5c of the Active Affine Shape Recovery Strategy), and (2) determining at which viewpoint the distance between the projections of two virtual points is maximized (Step 6). Since errors in image measurements can influence these operations, they should be taken into consideration. Below we consider the effect of measurement errors in the extremum-finding operation of Step 5c. Specifically, we show that the position of the virtual points defined by the strategy influences the strategy's sensitivity to noise. We also show how sensitivity can be improved through an iterative process that modifies the positions of the virtual points.

Suppose  $v$  is a virtual point on epipolar plane  $E$ , and suppose we want to compute the affine coordinates of a point in  $P_v$ . Let  $r(t)$  be the projection of  $v$  along  $\xi(t)$ , and let  $q(t)$  be an occluding contour point for which the image separation,  $\|q(t) - r(t)\|$ , exhibits a local extremum at  $t = t^*$ . We consider effects of corrupting  $\|q(t) - r(t)\|$  with a measurement error  $e(t)$ . Theorem 2 gives a second-order approximation to the reconstruction error,  $\delta$ :

**Theorem 2** Let  $\delta$  be the distance of  $S \cap E$  from the reconstructed point. If  $\|q(t) - r(t)\| + e(t)$  is locally extremal at  $\xi(t_{err})$  and  $|t^* - t_{err}|$  is small,  $\delta$  is given by

$$\delta \approx e(t_{err}) - 1 + \sqrt{1 + \frac{D}{R} \sin^2 \theta} \quad (2)$$

where  $R$  is the radius of curvature of  $S \cap E$  at the point  $p \in P_v$  corresponding to  $\|q(t^*) - r(t^*)\|$ ;  $D$  is the distance between  $v$  and the center of the osculating circle [21] of  $S \cap E$  at  $p$ ; and  $\theta$  is the angle between  $\xi(t_{err})$  and  $\xi(t^*)$ .

Theorem 2 tells us that the distance between a virtual point and the surface can have a dramatic effect on the quality of the reconstruction results: When  $v$  is near the center of the osculating circle of  $S \cap P$  at  $p$ , the reconstruction error is only linearly dependent on the error  $e(t)$ . On the other hand, if  $v$ 's distance from the center of the osculating circle is large, the reconstruction error is dominated by this distance (Figure 4). Resistance to noise can therefore be improved by selecting a point  $v$  close to the center of  $p$ 's osculating circle.

So suppose the reconstructed point is  $p_{rec}$  and let  $O$  be the center of the osculating circle at  $p$ . We use a method which relies on the fact that when  $\xi(t_{err})$  is near  $\xi(t^*)$ , the position of  $O$  relative to  $v$  and  $p_{rec}$  can be determined from the behavior of  $\|q(t) - r(t)\|$ . This quantity exhibits a local minimum at  $t = t^*$  if and only if  $v$  lies between  $p_{rec}$  and  $O$ , and exhibits a local maximum if and only if  $O$  lies between  $p_{rec}$  and  $v$ . To decrease the distance between points  $O$  and  $v$ , we perform a binary search on the line through  $v$  and the reconstructed point. Depending on the extremum type of  $\|q(t) - r(t)\| + e(t)$ ,  $v$  is repositioned either mid-way between  $p_{rec}$  and its current position, or at twice its current distance from  $p_{rec}$ . Both operations can be performed through linear combinations of the affine coordinates of  $p_{rec}$  and  $v$ .

## 7 Experimental Results

To demonstrate the effectiveness of our affine shape recovery approach we have performed preliminary experiments using both simulated and real data. Simulations were used to investigate the method's behavior for different shapes of the object's cross-section and in the presence of noise. We considered cylindrical objects with elliptical cross-sections.<sup>4</sup> Simulated data were generated by (1) randomly selecting four points on the object to be the affine basis points, and (2) computing the occluding contour and the projection of the affine basis points at 100 equally-spaced viewpoints corresponding to a rotation of  $2\pi$  radians about the object's axis. The occluding contour and the projected affine basis points were the only inputs to the system. Measurement noise was incorporated by perturbing the image coordinates of the affine basis points and the occluding contour points at every viewpoint by zero-mean, normally-distributed random values.

Tables 1 and 2 summarize the simulation results. Table 1 shows the reconstruction errors when the position of the virtual points is fixed. The table shows that noise sensitivity increases considerably when the ratio of the major and minor axes of the ellipse decrease, exactly as predicted by Theorem

<sup>4</sup>The objects' axis was kept perpendicular to the plane of the cross-section and parallel to the image plane.

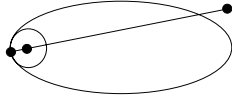


Figure 4: Recovering elliptical cross-sections. When a virtual point  $v$  is positioned close to a high-curvature segment of the ellipse, the radius of curvature at  $p \in P_v$  is much smaller than the distance between  $v$  and the center  $O$  of the osculating circle at  $p$ . Theorem 2 tells us that this distance will dominate the reconstruction error when measurement noise is present.

$\rho$	Noise variance				
	0%	0.2%	0.4%	0.6%	0.8%
0.5	0.8330	2.0414	15.9003	1.4364	2.7340
0.37	1.5683	1.8512	2.9530	2.3226	6.3211
0.25	2.5076	3.8449	5.3562	5.3207	6.8293
0.1	3.8645	4.3084	6.2938	13.4600	26.3474

Table 1: Reconstruction errors with no virtual point position updates. Errors are computed by averaging over 20 runs of the system the mean reconstruction error for 10 points reconstructed on a single epipolar plane. Errors are shown for different values of the ratio  $\rho$  of the cross-section’s minor and major axes. Noise variance is given as a percentage of the cross-section’s major axis. This axis was kept constant throughout the experiments.

2 (Figure 4). Table 2 shows that these errors can be reduced by iteratively repositioning the virtual points defined in Step 2 of the Active Affine Shape Recovery Strategy according to the method outlined in Section 6. This repositioning operation is particularly effective when the curvature of the surface varies considerably along the cross-section (i.e., when  $\rho$  is small).

Figure 5 shows the results of applying our strategy to a real scene. The object was placed on a platform mounted on the end-effector of a PUMA 700 manipulator. Four easily-identifiable markers were rigidly attached to this platform to define the affine basis. The occluding contour edges and the four affine basis points were then automatically tracked while the object was rotated about a vertical axis. A total

$\rho$	Noise variance				
	0%	0.2%	0.4%	0.6%	0.8%
0.5	0.4807	0.8022	1.2538	1.4810	1.7735
0.37	1.0403	1.2550	2.5328	4.0618	1.4557
0.25	1.8557	1.9900	2.5822	3.6199	4.3595
0.1	2.4174	6.8125	3.6932	4.5535	8.4308

Table 2: Reconstruction errors after two iterations of the virtual point repositioning method of Section 6.

of 256 images were obtained corresponding to a rotation of approximately  $2\pi$  radians. No information about the object’s motion or about the camera’s calibration parameters was used. The position of the virtual points was updated using two iterations of the method outlined in Section 6.

The reconstruction results can be evaluated by re-projecting reconstructed cross-sections along different viewing directions. Figures 5(b),(c) show that the object’s circular cross-section was correctly reconstructed. The reconstructed points do not lie on a single plane, however, as evidenced by the points’ re-projection in Figure 5(d). Such errors can be reduced by incorporating a constraint in the reconstruction of individual points that enforces the planarity of the cross-section. In addition to simply recovering the positions of points on the cross-section, Figure 5(d) shows that our shape recovery strategy can be used to obtain surface curvature information: The projection of the center of the osculating circle at different points on the cross-section can be localized by iteratively updating the position of virtual points according to the method of Section 6.

## 8 Concluding Remarks

We are currently considering a number of extensions to the basic approach described in this paper. These include (1) studying how robustness to noise can be improved by constraining the reconstruction of nearby points, (2) extending our active shape recovery approach to the case of perspective projection, and (3) studying how accurate planar motions on the tangent plane of arbitrary visible rim points can be achieved with an uncalibrated hand-eye system.

## References

- [1] J. L. Mundy and A. Zisserman, eds., *Geometric Invariance in Computer Vision*. MIT Press, 1992.
- [2] P. A. Beardsley, A. Zisserman, and D. W. Murray, “Navigation using affine structure from motion,” in *Proc. 3rd European Conf. on Computer Vision*, pp. 85–96, 1994.
- [3] O. D. Faugeras, “What can be seen in three dimensions with an uncalibrated stereo rig?,” in *Proc. 2nd European Conf. on Computer Vision*, pp. 563–578, 1992.
- [4] D. Weinshall and C. Tomasi, “Linear and incremental acquisition of invariant shape models from image sequences,” in *Proc. 4th Int. Conf. on Computer Vision*, pp. 675–682, 1993.
- [5] A. Shashua, “Projective structure from uncalibrated images: Structure from motion and recognition,” *IEEE Pattern Anal. Machine Intell.*, vol. 16, no. 8, pp. 778–790, 1994.
- [6] C. A. Rothwell, D. A. Forsyth, A. Zisserman, and J. L. Mundy, “Extracting projective structure from single perspective views of 3d point sets,” in *Proc. 4th Int. Conf. on Computer Vision*, pp. 573–582, 1993.
- [7] J. J. Koenderink and A. J. van Doorn, “Affine structure from motion,” *J. Opt. Soc. Am.*, vol. A, no. 2, pp. 377–385, 1991.
- [8] Y. Aloimonos, “Purposeful and qualitative active vision,” in *Proc. Int. Conf. on Pattern Recognition*, pp. 346–360, 1990.

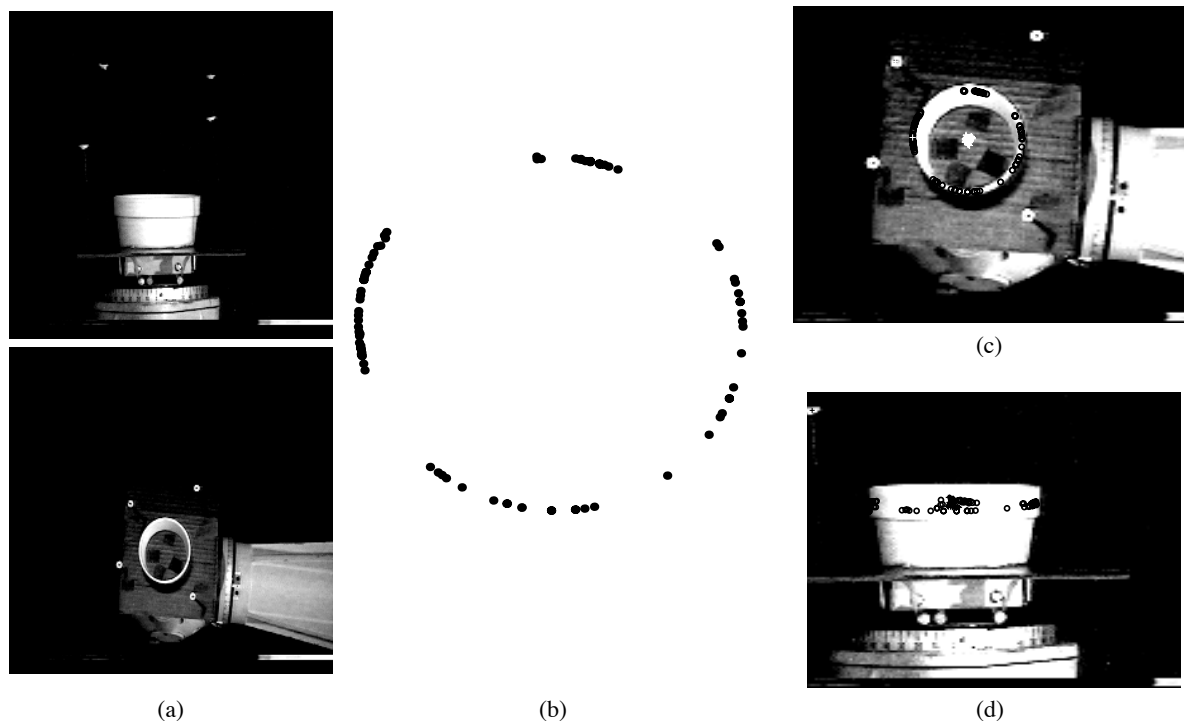


Figure 5: (a) Two views of a cup. Four markers (bright dots) were mounted on the platform supporting the cup to define the affine basis. (b) Reconstruction results. Points are re-projected along the direction corresponding to the image in (c). (c) Reconstructed surface points, re-projected along a near-top view of the object and overlaid with the object's image. Also overlaid are the re-projected virtual points used for reconstruction (white points near cup's axis of revolution). The position of these points was updated through two iterations of the virtual point re-positioning method of Section 6. (d) View of the reconstructed points along a direction in the epipolar plane.

- [9] D. H. Ballard, "Animate vision," *Artificial Intelligence*, vol. 48, pp. 57–86, 1991.
- [10] J. J. Koenderink, *Solid Shape*. MIT Press, 1990.
- [11] R. Cipolla and A. Blake, "Surface shape from the deformation of apparent contours," *Int. J. Computer Vision*, vol. 9, no. 2, pp. 83–112, 1992.
- [12] K. N. Kutulakos and C. R. Dyer, "Occluding contour detection using affine invariants and purposive viewpoint control," in *Proc. Computer Vision and Pattern Recognition*, pp. 323–330, 1994.
- [13] P. Giblin and R. Weiss, "Reconstruction of surfaces from profiles," in *Proc. 1st Int. Conf. on Computer Vision*, pp. 136–144, 1987.
- [14] R. Vaillant and O. D. Faugeras, "Using extremal boundaries for 3-d object modeling," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, no. 2, pp. 157–173, 1992.
- [15] R. Szeliski and R. Weiss, "Robust shape recovery from occluding contours using a linear smoother," in *Real-time Computer Vision* (C. M. Brown and D. Terzopoulos, eds.), pp. 141–165, Cambridge University Press, 1994.
- [16] N. Ahuja and A. L. Abbott, "Active stereo: Integrating disparity, vergence, focus, aperture, and calibration for surface estimation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, no. 10, pp. 1007–1029, 1993.
- [17] J. Ponce, D. Chelberg, and W. B. Mann, "Invariant properties of straight homogeneous generalized cylinders and their contours," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 9, pp. 951–966, 1989.
- [18] D. A. Forsyth, "Recognizing algebraic surfaces from their outlines," in *Proc. 4th Int. Conf. on Computer Vision*, pp. 476–480, 1993.
- [19] L. S. Shapiro, A. Zisserman, and M. Brady, "Motion from point matches using affine epipolar geometry," in *Proc. 3rd European Conf. on Computer Vision*, pp. 73–84, 1994.
- [20] K. N. Kutulakos, "Affine surface reconstruction by purposive viewpoint control," Tech. Rep. 581, Department of Computer Sciences, University of Rochester, 1995.
- [21] M. P. D. Carmo, *Differential Geometry of Curves and Surfaces*. Englewood Cliffs, NJ: Prentice-Hall Inc., 1976.