To appear in: Proc. CVPR'94, Seattle, WA

Occluding Contour Detection Using Affine Invariants and Purposive Viewpoint Control

Kiriakos N. Kutulakos
kyros@cs.wisc.edu

Charles R. Dyer
dyer@cs.wisc.edu

Computer Sciences Department
University of Wisconsin
Madison, WI 53706

Abstract

We present an approach for identifying the occluding contour and determining its sidedness using an active (i.e., moving) observer. It is based on the non-stationarity property *of the visible rim: When the observer's viewpoint is changed, the visible rim is a collection of curves that "slide," rigidly or non-rigidly, over the surface. We show that the observer can deterministically choose three views on the tangent plane of selected surface points to distinguish such curves from stationary surface curves (i.e., surface markings). Our approach demonstrates that the occluding contour can be identified* directly*, i.e., without first computing surface shape (distance and curvature).*

1 Introduction

This paper presents a new approach employing a mobile observer that *purposefully* controls its motion in order to identify the image curves corresponding to the occluding contour of a surface. The approach is motivated by the purposive [1] and animate [2] vision paradigms, and combines elements from the study of the occluding contour of smooth surfaces [3–6] and previous work on non-metric scene reconstruction [7–9]. We show that by using at least three distinct views of the surface we can (1) identify the occluding contour curves in an image, and (2) determine on which side of these curves the surface lies (i.e., the "sidedness" of the contour). The only requirements of this approach are that (1) a correspondence can be established in the three views for the image curves whose identity is sought (i.e., "surface marking" or "occluding contour"), (2) at least four image curve points with parallel tangents can be identified, and (3) the motion of the observer has a specific relationship with the geometry of the viewed scene. We show that this relationship is characterized by specific image-computable quantities and enables an invariant-based analysis [8]. Our approach does not require knowledge of observer velocities or object models, uses a world-centered coordinate frame, and can incorporate multiple views to improve stability.

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 and is a rich source of both quantitative and qualitative surface information [3, 4]. Under continuous motion of the observer the visible rim possesses two important properties: (1) The curves comprising the visible rim "slide" rigidly or non-rigidly over the surface, possibly changing their connectivity, and (2) the deformation of the visible rim's projection, the occluding contour, uniquely determines the shape (i.e., curvature) of the surface regions over which the visible rim slides [3, 6].

Most previous work on identifying the occluding contour of smooth surfaces focused on the second property: Because occluding contour deformations uniquely determine shape, one can check whether the deformations of an image curve determine the shape of a smooth surface. Since the deformation of the projection of a surface marking corresponds to a surface of infinite curvature coinciding with the marking itself [3, 5], this test is sufficient for identifying the occluding contour curves in a sequence of images. Unfortunately, the test involves comparing the speeds and accelerations of the curves in the image sequence to the speed and acceleration of the observer, and hence, requires accurate measurement of these quantities (or an accurately calibrated stereo system [5]). In fact, it has been shown that (1) the robustness of the surface reconstruction process itself can be greatly improved by *a priori* distinguishing between the occluding contour and the surface markings [3], and (2) specialized viewpoint control strategies can be used to simplify the recovery of surface curvature information once the occluding contour has been identified [10]. Another drawback of recovering shape before identifying the occluding contour is that when the occluding contour is used only for describing the scene qualitatively, e.g., for model indexing or for grasping an object, accurate shape information and knowledge of viewer motion is not necessary.

In this paper we show how the occluding contour can be identified *directly*, i.e., without first computing surface shape (distance and curvature). We achieve this, as in [11], by exploiting the first property of the visible rim: Unlike the visible rim curves, surface markings are *stationary*, i.e., their position on the surface is fixed and viewpoint-independent. Instead of attempting to accurately measure the speeds and accelerations of the visible rim curves over the surface, we utilize the stationarity property of the sur-

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

face markings; by determining which curves in the image are projections of stationary curves we factor out the need for recovering metric properties about the surface or the observer's motion.[1]

Our approach is based on a simple observation: Suppose a curve on the visible rim coincides with a surface marking. Because the position of the visible rim curve depends on viewpoint while the position of the surface marking does not, the coincidence relationship between the projections of the two curves will not be preserved when the observer's viewpoint changes. So, if we are able to compute how a visible surface curve would project at other viewpoints assuming it is stationary, we can simply change viewpoint and compare its projection with the one predicted under the stationarity assumption. When the predicted projection of a curve does not coincide with the actual one, the curve must be on the visible rim.

The crucial issue one must address to exploit this observation is how to predict the projection of a surface curve under the stationarity assumption. Previous work on predicting novel views of a three-dimensional object either assumed the existence of a collection of "model" views in which the projections of either the surface markings or the visible rim curves have been identified [12], or the existence of a small number of easily-identifiable point features on the surface of an unknown object that could be matched across frames [11, 13]. Here we present a detailed geometrical analysis of this prediction problem that (1) shows how to distinguish stationary from non-stationary surface curves using an active observer and discusses under what conditions this discrimination can be achieved, (2) shows how this ability can be used to detect the occluding contour as well as the contour's sidedness when no surface markings have been *a priori* identified, and (3) takes into account errors in image measurements.

The basic assumption used in previous approaches for identifying the occluding contour and for predicting novel views of an object was that the observer motion was *arbitrary*. This means that the motion of the observer between any two views of the scene is not related in any way to the geometry of the scene. This is a reasonable approach, however, only when the observer cannot control its viewpoint. When the observer's motion can be controlled, the choice of viewpoint(s) does not have to be arbitrary.

The significance of our method lies in the use of purposive observer motion to achieve and maintain specific geometric relationships with the viewed surface in order to distinguish the occluding contour from the projections of surface markings. We show how we can use an active observer to obtain a collection of views in which the image of a stationary surface curve can be accurately predicted, and then use this predictive power to classify image curves either as occluding contour curves, or as projections of stationary surface curves. Throughout this paper we assume that each image has been processed to extract a collection of curves.

[1] When the position of the light source and the object is fixed, image curves corresponding to shadows are also stationary. We do not distinguish here between such image curves and the projections of surface markings.

2 Viewing Geometry

Let S be a smooth, oriented surface in $\Re^3$, viewed under orthographic projection along a viewing direction ξ. The *visible rim* of S along ξ is the set of the visible surface points p for which the tangent plane, $T_p(S)$, contains a line parallel to ξ. The *occluding contour* is the projection of the visible rim on the image plane. The occluding contour and the visible rim each consist of a collection of open and closed smooth curves for almost all viewpoints of the observer [4].

The shape and position of the visible rim depends on the shape of S and the observer's viewpoint. This is the fundamental difference between visible rim curves and surface markings; the surface position of the curves corresponding to such markings is independent of the observer's viewpoint. We qualitatively characterize the difference between these two types of curves with the notion of *stationarity*:

Definition 1 A surface curve is *stationary* if its position on the surface does not change when the observer's viewpoint changes. It is *non-stationary* if its position is viewpoint-dependent.

Our goal is to exploit the non-stationarity of the visible rim to identify the occluding contour. This operation requires a correspondence between points on the surface at different viewpoints, and between points in images. We use the epipolar plane correspondences [3, 5, 6] for this purpose.

In particular, suppose the observer's viewing direction changes according to $\xi(t)$ on a motion plane with normal $\mathbf{n_E}$. This motion defines a family of planes parallel to the observer's motion plane called the *epipolar planes*. If $B(t_0)$ is an image curve at time t_0, the epipolar plane correspondence matches a point $q \in B(t_0)$ to the intersection of $B(t_0 + \delta t)$ with the epipolar plane through q. This correspondence induces a correspondence between points belonging to the surface curve $\beta(t)$ projecting to $B(t)$ (Figure 1).

Figures 1(a),(b) make explicit the non-stationarity property characterizing the visible rim curves: When the tangent of an occluding contour curve at a point q does not belong to an epipolar plane, we can think of the point $q(t)$ corresponding to q at time t as being the projection of a point $p(t)$ that "moves" on the intersection of the surface with the epipolar plane through q. We call such points $p(t)$ *non-stationary*.

The key property of the epipolar correspondences used in our approach is derived from the geometry of Figure 1(c): Even though, in general, we cannot determine which points in the image are projections of stationary points, we can *force* the stationarity of specific points by appropriately controlling viewpoint; the observer simply needs to move in a plane parallel to the tangent plane at those points.

Tangential Motion Property: If a point q on an image curve has its tangent parallel to the observer's motion plane, and epipolar plane correspondences have been established for q across frames, all points corresponding to q must be projections of the *same* surface point. This property does not depend on whether or not the image curve containing q is the projection of a surface marking.

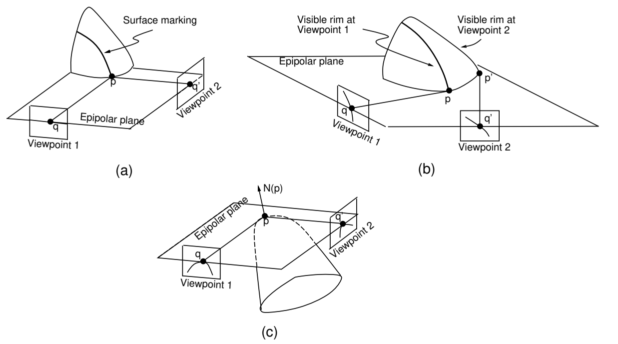

Figure 1: Point correspondences induced by the epipolar geometry. Point q' is the point corresponding to q. (a) When q is the projection of a point p on a surface marking at Viewpoint 1, q' is the projection of p at Viewpoint 2. (b) When p is on the visible rim and the tangent of the image curve at q does not belong to an epipolar plane, $T_p(S)$ is not an epipolar plane and q' is the projection of p', not p. (c) When the tangent of the image curve at q belongs to an epipolar plane, $T_p(S)$ is an epipolar plane and both q and q' are projections of p [10, 14].

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 $\xi(t)$.

2.1 Affine-Invariant Representations

A basic step in our method for determining the non-stationarity of the visible rim is that of *re-projection* [12,13]: Given the projections of a collection of 3D points along a sequence of viewing directions, compute the projection of those points along a viewing direction that is not contained in the sequence. Affine-invariant representations are important because they allow us to re-project points without knowing the observer's motion and without recovering any metric properties of the corresponding surface points.[2]

Let $p_1, \dots, p_n \in \Re^3, n \geq 4$, be a collection of points at least four of which are not coplanar. An *affine-invariant representation* of those points consists of three components: The *origin*, p_o, which is one of the points $p_i, i = 1 \dots, n$; the *affine basis points*, $p_{b_1}, p_{b_2}, p_{b_3}$, which are three points from the collection that are not coplanar with the origin; and the *affine coordinates* of the points p_i, which are the coordinates of $p_i - p_o$ with respect to the affine basis vectors $\mathbf{b}_j = p_{b_j} - p_o, j = 1, 2, 3$. To re-project, we use the following two properties of affine-invariant representations:[3]

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

$$(1) \quad \begin{bmatrix} u_p^m \\ v_p^m \end{bmatrix} = \begin{bmatrix} u_{\mathbf{b}_1}^m & u_{\mathbf{b}_2}^m & u_{\mathbf{b}_3}^m \\ v_{\mathbf{b}_1}^m & v_{\mathbf{b}_2}^m & v_{\mathbf{b}_3}^m \end{bmatrix} \mathbf{A_p} - \begin{bmatrix} u_{p1}^m \\ v_{p1}^m \end{bmatrix}$$

where (u_p^m, v_p^m) is the projection of p and $\mathbf{A_p}$ is the vector collecting the affine coordinates of p.

Property 2 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.

3 Detecting Point Non-Stationarity

The previous section showed that a fundamental property distinguishing the visible rim from surface markings is *non-stationarity*: Except for a finite (and, in practice, small) collection of points that depends on the observer's motion plane, the points on the visible rim are non-stationary. Hence, we can identify the occluding contour curves by using an active observer to determine which image curves contain projections of non-stationary points.

Our approach is based on a theorem that allows us to distinguish between the projection of stationary and non-stationary points by exploiting a few simple properties of smooth surfaces and their occluding contours. In particular, suppose the observer's viewing direction changes according to a smooth function $\xi(t)$, let $p(0)$ be the surface point projecting to $q(0)$, and let E be the epipolar plane through $p(0)$. Theorem 1 characterizes the distance between $q(t)$ and the projection, $\hat{q}(t)$, of $p(0)$ along $\xi(t)$:

Theorem 1 Points $p(t), q(t)$, and $\hat{q}(t)$ have three properties:

(1) Suppose that the curves in $S \cap E$ are parameterized so that their curvature is positive when their normal is toward the surface interior. Then, $p(t)$ is contained in the maximal, connected, convex subset of $S \cap E$ that contains $p(0)$.

(2) Let λ be the smooth curve corresponding to the set in (1). The distance between $q(t)$ and $\hat{q}(t)$ is non-zero only if $p(0)$ is non-stationary, and is equal to $|\,[p(t) - p(0)] \cdot \mathbf{n}(\mathbf{p}(\mathbf{t}))|$, where $\mathbf{n}(\mathbf{p}(\mathbf{t}))$ is the normal of λ at $p(t)$.

(3) If $p(0)$ is non-stationary, $\|q(t) - \hat{q}(t)\|$ is zero along at most three viewing directions on E, including $\xi(0)$.

Proofs can be found in [16]. Theorem 1 motivates the use of a prediction-verification scheme for determining the stationarity of a point $p(0)$: If we are able to compute $p(0)$'s projection along any viewing direction on the observer's motion plane, we can assume that $p(0)$ is stationary and check the validity of that assumption. $p(0)$'s can be verified by moving to a new viewing direction $\xi(t)$ and then verifying that $\hat{q}(t)$, i.e., the *predicted* position of $q(t)$ under the stationarity assumption, coincides with $q(t)$ in the new image. If $q(t)$ and $\hat{q}(t)$ do not coincide, $p(0)$ must be non-stationary. Intuitively, if non-zero, the distance $\|q(t) - \hat{q}(t)\|$ measures the "degree" of non-stationarity of $p(0)$.

The following corollary to Theorem 1 goes a step further, showing that the position of $\hat{q}(t)$ can serve as a qualitative indicator of the occluding contour's sidedness (Figure 2):

[2] The use of affine invariants for detecting the occluding contour was also suggested in [11].

[3] See [7, 8, 15] for the related theory, details on these properties, and for algorithms to compute the affine representation of a collection of points.

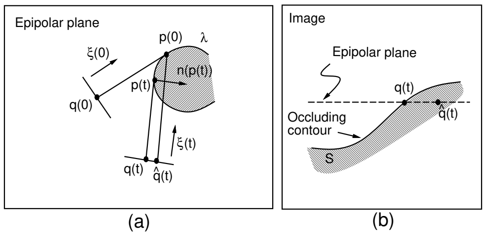

Figure 2: Determining the occluding contour's sidedness. (a) Since λ is convex, the vector $\hat{q}(t) - q(t)$ points in the direction of λ's normal, $\mathbf{n}(\mathbf{p}(\mathbf{t}))$, at $p(t)$. The corresponding image is shown in (b). The epipolar plane is always viewed "edge-on", and hence it projects to a line in the image.

Corollary 1 If $p(0)$ is non-stationary and $\xi(t_0)$ is such that $\|q(t) - \hat{q}(t)\| \neq 0$ for $0 < t \leq t_0$, the surface must lie on the side of $q(t_0)$ containing $\hat{q}(t_0)$.

Theorem 1 and Corollary 1 tell us that even an arbitrarily small change in viewing direction on E will force $\|q(t) - \hat{q}(t)\|$ to become non-zero. Even though, in theory, this should be sufficient for identifying the non-stationary points in an image, in practice we must allow for errors in image measurements and the inability to measure arbitrarily small distances in an image. The next corollary to Theorem 1 characterizes the effectiveness of this prediction-verification approach by specifying how large $\|q(t) - \hat{q}(t)\|$ can be:

Corollary 2 The function $f(t) = \|q(t) - \hat{q}(t)\|$ is maximized when $[p(t) - p(0)] \cdot \xi(t) = 0$, i.e., when $p(t)$ is at a local maximum distance from $p(0)$. (1) If at least one such point exists on λ and that point is not one of its endpoints, then there is a point $p(t^*)$ corresponding to a maximum of f, such that

$$f(t^*) \geq \frac{|\sin\phi|}{\kappa_{max}} \tag{2}$$

where κ_{max} is the maximum absolute principal curvature of the surface, and ϕ is the minimum angle between E's normal and the surface normal along λ. (2) If $f(t)$ is maximized at λ's endpoints,

$$f(t^*) \geq \min\{D(p_{mid}, l_1), D(p_{mid}, l_2)\} \tag{3}$$

where p_1, p_2 are the two endpoints of λ; l_1, l_2 are the tangent lines at p_1, p_2, respectively; p_{mid} is the midpoint of λ; and $D(p, l)$ is the distance from point p to line l.

4 Active Occluding Contour Detection

The goal of our approach is to control the observer's viewpoint so that the identity (visible rim or surface marking) of the surface curves projecting to the image can be determined. We do this by assuming the curves projecting to the image are stationary, selecting a motion plane, and then controlling viewpoint on that plane to verify the stationarity assumption. This process involves verifying that the actual projection of the surface curves coincides with that predicted under the assumption that they are stationary. To completely specify our occluding contour detection strategy we therefore need to answer three questions: (1) How to select the motion plane, (2) how to control viewing direction on that plane, and (3) how to predict the projection of the selected curves from viewpoints on that plane under the stationarity assumption? Below we first outline the main ideas, and in Sections 5 and 6 present more details.

In particular, suppose that the observer's motion plane has been selected and that we want to control the observer's viewing direction to determine the non-stationarity of a single point $p(0)$ on the visible rim at the observer's initial viewpoint. Let $q(0)$ be $p(0)$'s projection. Theorem 1 tells us that there are viewing directions on the observer's motion plane from which the non-stationarity of $p(0)$ can be established by verifying that $p(t)$'s projection does not coincide with that of $p(0)$. Since the distance between the two projections can be arbitrarily small depending on the observer's viewpoint, we use a strategy to reach viewpoints for which this distance is large. Corollary 2 shows that in most cases, the maximum distance can be characterized by only two parameters, one of which depends only on the intrinsic surface geometry. When the observer's goal is to identify the type of each curve in a collection of curves in an image, this process can be applied simultaneously to all points on the selected curves. These considerations lead to the following strategy for identifying the occluding contour:

Active Occluding Contour Detection Strategy

Step 1: Let $B_1, \ldots, B_n$ be the image curves whose identities are sought.

Step 2: Select the motion plane E.

Step 3: Perform a small clockwise viewing direction change.

Step 4: For every point $q(0)$ on $B_1(0), \ldots, B_n(0)$ for which a correspondence $q(t)$ can be established along the current viewing direction,

- **a.** compute $\hat{q}(t)$,
- **b.** compute the distance $\delta(q(t)) = \|q(t) - \hat{q}(t)\|$.

Step 5: Repeat Steps 3-5 until either the initial viewing direction is reached, or no correspondences can be established for curves $B_1(0), \ldots, B_n(0)$.

Step 6: If the initial viewing direction is not reached, repeat Steps 4-5 while changing viewing direction in a counter-clockwise fashion.

Step 7: Given the computed distances $\delta(q(t)), 0 \leq t \leq T$, for each point $q(0)$ on the curves $B_1, \ldots, B_n$, label these curves as either type surface markings or type occluding contour.

The next section shows how we can use affine invariant representations to perform Step 4, and how that step also determines the observer's motion plane. Section 6 then discusses the curve classification process in Step 7, which takes into account noise in image measurements.

5 A Prediction Mechanism for Detecting Non-Stationarity

The occluding contour detection strategy described in the previous section made an important assumption: Given a point $p(0)$ on the surface, the observer can compute $p(0)$'s projection along arbitrary viewing directions on its motion plane. In general, this computation cannot be performed unless the three-dimensional coordinates of that point are known. Here we show that by exploiting the Tangential Motion Property (Section 2) to select the observer's motion plane we can formulate this computation as a re-projection; this allows the assumption to be slightly relaxed without affecting the correctness of the strategy given in Section 4.

Suppose we want to determine whether or not a point $p(t_1)$ projecting to $q(t_1)$ is non-stationary. Theorem 1 says that if we select a plane through $p(t_1)$ that is not tangent to $p(t_1)$ and change our viewing direction on that plane to $\xi(t)$, we can determine that $p(t_1)$ is non-stationary by verifying that the predicted projection, $\hat{q}(t)$, of $p(t)$ along $\xi(t)$ does not coincide with the measured one, i.e., $q(t)$. To exploit this theorem without having to explicitly compute $\hat{q}(t)$ we simply note that the exact distance between $\hat{q}(t)$ and $q(t)$ is not important, as long as it is non-zero. We use the following result, which is similar to Theorem 1:

Theorem 2 Let $p(t_1)$ be a surface point and let $q(t_1)$ be its projection along $\xi(t_1)$. Let $q(t_2)$ be the point corresponding to $q(t_1)$ along a viewing direction $\xi(t_2)$ on the observer's motion plane, E. If (1) the projections, $r_1(t_1), \ldots, r_4(t_1)$, of four stationary and non-coplanar surface points can be identified in the image along $\xi(t_1)$, and (2) $\mathbf{A}$ is the affine coordinate vector computed using Property 2 and Eq. 1 with $(u_p^m v_p^m) = q(t_m), m = 1, 2$, and $(u_{\mathbf{b}_j}^m v_{\mathbf{b}_j}^m) = r_j(t_m) - r_1(t_m), j = 2, 3, 4$, then the point $\tilde{q}(t)$ defined by

$$\tilde{q}(t) = [r_2(t) - r_1(t) \;\; r_3(t) - r_1(t) \;\; r_4(t) - r_1(t)]\mathbf{A} \tag{4}$$

has the following two properties:

1. Let λ be defined as in Theorem 1, and let $\tilde{p}$ be the 3D point with affine coordinates $\mathbf{A}$. The distance between $q(t)$ and $\tilde{q}(t)$ is non-zero only if $p(t_1)$ is non-stationary, and is equal to $|[p(t) - \tilde{p}] \cdot \mathbf{n}(\mathbf{p}(\mathbf{t}))|$, where $\mathbf{n}(\mathbf{p}(\mathbf{t}))$ is the normal of λ at $p(t)$.
2. If $p(t_1)$ is non-stationary, $\|q(t) - \tilde{q}(t)\|$ is zero along at most four viewing directions on E.

Theorem 2 implies that if we can identify the projections of four non-coplanar stationary points in the image, we can replace the computation of $\hat{q}(t)$ in Step 4a of the Active Occluding Contour Detection Strategy by computing $\tilde{q}(t)$ without compromising the correctness of the strategy. The first property of $\tilde{q}(t)$ ensures that by computing $\tilde{q}(t)$ instead of $\hat{q}(t)$ we do not introduce any "false positives" in our occluding contour detection process. The second property shows that the distance $\|q(t) - \tilde{q}(t)\|$ behaves almost identically to the "true" distance $\|q(t) - \hat{q}(t)\|$, allowing us to use the Active Occluding Contour Detection Strategy to identify the occluding contour. The following corollary, which generalizes Corollaries 1 and 2, makes this explicit:

Corollary 3 (1) The function $\tilde{f}(t) = \|q(t) - \tilde{q}(t)\|$ is maximized when $[p(t) - \tilde{p}] \cdot \xi(t) = 0$, i.e., when $p(t)$ is a local maximum distance from $\tilde{p}$. If at least one such point exists on λ and that point is not one of its endpoints, then there is a point $p(t^*)$ corresponding to a local maximum of $\tilde{f}$, such that

$$\tilde{f}(t^*) \geq \frac{|\sin\phi|}{\kappa_{max}} \tag{5}$$

where κ_{max} is the maximum absolute principal curvature of the surface, and ϕ is the minimum angle between E's normal and the surface normal along λ. (2) Let $\xi(t^*)$ be the viewing direction maximizing $\tilde{f}$. If $p(0)$ is non-stationary, the surface must lie on the side of $q(t^*)$ containing $\tilde{q}(t)$.

Corollary 3 shows that the only major impact of computing $\tilde{q}(t)$ instead of $\hat{q}(t)$ is in the determination of the contour's sidedness. Intuitively, the contour's sidedness can be correctly determined only along viewing directions at or close to the global maximum of $\|q(t) - \tilde{q}(t)\|$. To use Theorem 2 and Corollary 3 we must be able to identify the projections of four stationary points in the image. In general, when the observer cannot control its motion this is impossible to achieve. However, when the observer can purposefully select its motion plane, the Tangential Motion Property can be exploited: If four points in the first image can be identified whose tangents are parallel, the motion plane defined by those tangents guarantees the points's stationarity. This solves the problem of predicting the projection, $\hat{q}(t)$, of $p(t)$ in Step 4a of the Active Occluding Contour Detection Strategy under the assumption that $p(0)$ is stationary.

6 Measurement Errors

The final step in the Active Occluding Contour Detection Strategy is the classification of image curves based on whether or not they contain projections of non-stationary points. This requires determining if a specific quantity, namely $\|q(t) - \tilde{q}(t)\|$, is zero or not. In practice, due to errors in image measurements, this quantity may not be zero even for projections of stationary points. It is therefore necessary to account for such errors in the classification process. We accomplish this in three ways: (1) We use many views to determine the non-stationarity of each point $p(0)$ projecting to the initial image by solving for $\mathbf{A}$ in Eq. 1 using least squares, (2) we estimate the probability that $p(0)$ is non-stationary under the assumption the image measurement errors have a zero-mean Gaussian distribution, and (3) to classify a curve, we estimate the above probability for all points projecting to that curve.

In particular, to classify point $p(0)$ as non-stationary, we compute the *squared Mahalanobis distance*, $\|\mathbf{q} - \tilde{\mathbf{q}}\|^2$, of the residual $\mathbf{q} - \tilde{\mathbf{q}}$ from $\mathbf{0}$, assume that this distance is corrupted by noise in the image measurements, and then estimate the probability that it is non-zero [17]. To simplify the error analysis we only allow for errors in the measurement of $\mathbf{q}$.[4] We also assume the errors in each element of

[4] A analysis taking into account errors in the projection of the basis

$\mathbf{q}$ are independent and have a Gaussian distribution with zero mean and variance σ. Under these assumptions, and when $p(0)$ is stationary, the distance $\frac{1}{\sigma^2}\|\mathbf{q}-\tilde{\mathbf{q}}\|^2$ follows a χ^2 distribution with three degrees of freedom [17, 18]. We now have the following criteria:

Point Classification Criterion: Classify $q(0)$ as an occluding contour point if

$$\Pr\left[\frac{1}{\sigma^2}\|\mathbf{q}-\tilde{\mathbf{q}}\|^2 > \delta \mid p(0) \text{ is stationary}\right] < m$$

where m is an *a priori* defined constant, δ is the computed value of $\frac{1}{\sigma^2}\|\mathbf{q}-\tilde{\mathbf{q}}\|^2$, and the probability is computed using a χ^2 distribution table.

Curve Classification Criterion: Classify curve B as an occluding contour curve if more than $M\%$ of B's points are classified as being on the occluding contour.

7 Experimental Results

To demonstrate the effectiveness of our active occluding contour detection approach we have performed preliminary experiments using both simulated and real data. Figure 3(a) shows one of the models used in our simulations, a polyhedral representation of a bottle. Four points were randomly selected from the polyhedron's vertices to be the affine basis points. The object was rotated a total of π radians about a vertical axis through the center of the bottle, and the projection of the affine basis points was computed for each of the frames. The generated images and the projections of the basis points were the only inputs to the system. The affine coordinate computations were performed using the first two frames of the sequence. This simulation corresponds to the best-case scenario for our strategy: Due to the model's symmetry, the occluding contour remains unchanged throughout the model's rotation, while the projection of a stationary surface curve varies with viewpoint. The simulation results are shown in Figures 3(b),(c). They show that the difference between the actual and predicted positions of the image curves can be dramatic, in this case almost equal to the radius of the bottle. Note also that the predicted position of the curve is on the side of the contour where the surface lies, correctly indicating the contour's sidedness.

Figures 3(d)-(f) and 4 show the results of applying our strategy to a real scene. The image sequence consisted of five frames, showing rotation of about $\pi/2$ radians. No information about the object's motion (apart from the direction of rotation) or the camera parameters was used. Viewing direction was changed in a horizontal plane perpendicular to the plane of the page. The system detected points with horizontal tangents and tracked them across the frames. The affine basis points were selected by minimizing the condition number of the affine basis matrix [15]. The point and curve classifications were performed as described in Section 6. We used $\sigma = 1$ and classified a curve as an occluding contour curve if more than 80% of its points had a probability greater than 90% of being non-stationary.

The results of the classification process for three of the image curves can be evaluated using the graphs in Figure 4. The first graph shows the variation of $\|\mathbf{q}-\tilde{\mathbf{q}}\|^2$ for the curve points with respect to their position on the curve. This quantity must be close to zero for stationary points and is the basic information used to determine a point's identity. Because we employed a very simple method for establishing point correspondences across frames, i.e., independently for each point on the curve, the second graph evaluates tracking performance. The graph plots the variance of the inter-frame matching distance for every point on the curve with respect to its position on the curve. Sharp peaks indicate tracking errors for the associated points. The third graph shows the input to the final stage of the curve classification process. It is a histogram of the number of points on the curve that have a given probability of being non-stationary. This information is used by the Curve Classification Criterion (Section 6) to determine the identity of the image curve. The mass under the histogram should be concentrated near 100% for non-stationary curves.

The "m" curve on the surface and the upper occluding contour curve (top of the head) were classified correctly, and the sidedness of the upper occluding contour curve was correctly determined. The approach's success is also indicated by the dramatic difference in their associated probability histograms. However, due to the small amount of rotation and the high surface curvature near the visible rim, the distance between the predicted and actual curves for both occluding contour curves in Figure 3(e) was not as great as in the case of the bottle. This resulted in a misclassification of the left curve corresponding to the right arm of the object. In addition, a considerable number of points were misclassified in the "m" curve. This was due to tracking errors, indicated by the strong correlation between errors in $\|\mathbf{q}-\tilde{\mathbf{q}}\|^2$ and sharp peaks in the variance of the inter-frame point matching distance (e.g., around point position 250). We do not currently use this correlation information to classify points. In the case of the upper occluding contour curve, a number of points were assigned low probability of being non-stationary. This is not an error; these points are near the top of the head, where the surface normal is close to that of the motion plane.

A number of observations can be made from these experiments. First, a curve may not contain points of only one type. For example, the upper occluding contour curve contains points on its bottom end that are not occluding contour points. However, a sharp drop in the prediction error could be used to decide where to segment such curves. Second, the results show that the approach is more effective with large motions. In effect, the approach factors out the need for measuring differential properties of the viewed surface by observing large-scale effects of viewpoint control on the surface's occluding contour. Of course, large motions require reliable curve tracking across many frames, as well as consistent tracking of points with tangents parallel to the observer's tangent plane. Tracking of such points will not,

points is also possible [5].

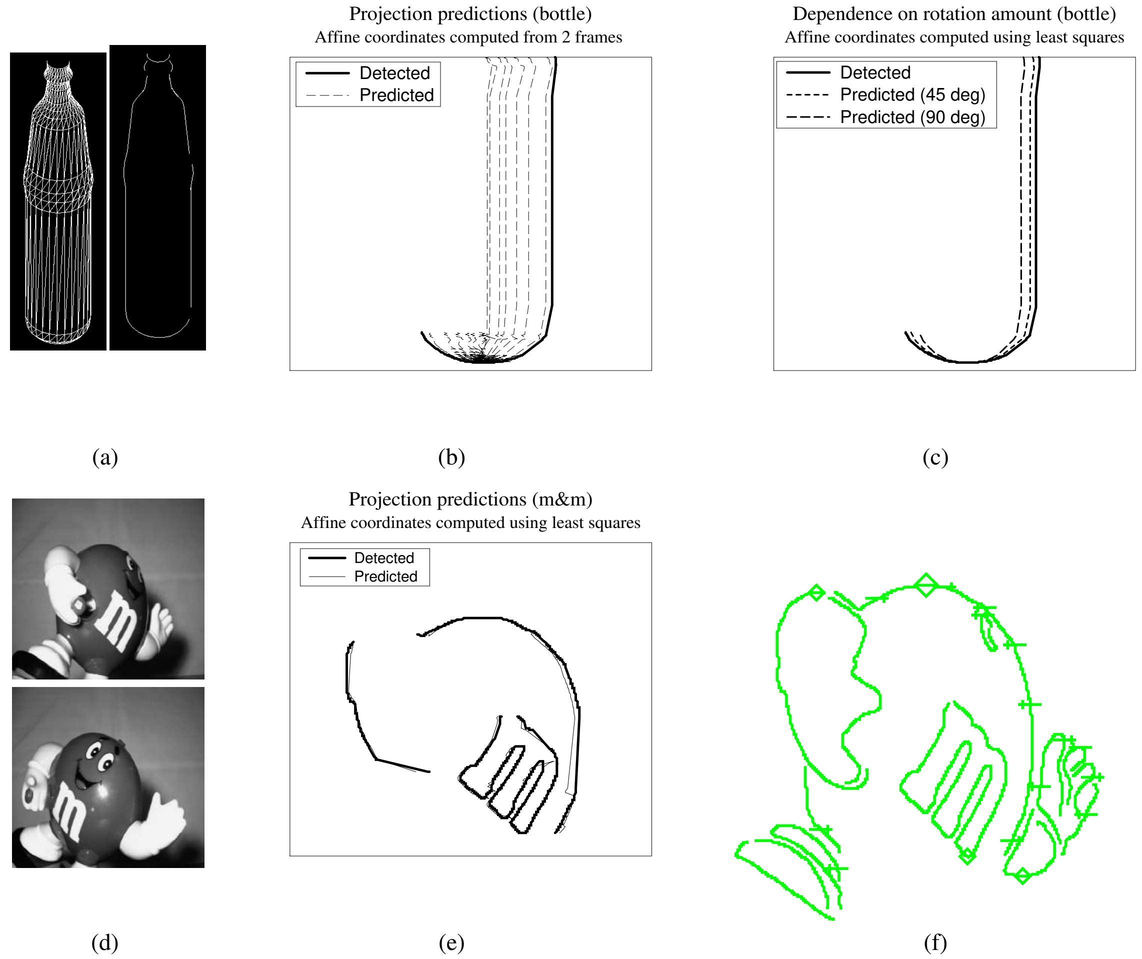

Figure 3: (a) Polyhedral model of a bottle and its visible rim. (b) Predictions for 10 frames overlayed with the detected curve which does not move throughout the rotation. (c) Predictions for the first frame when 10 frames (rotation of $\pi/4$ radians) and 20 frames (rotation of $\pi/2$ radians) are used for the least-square affine computations. Note that the distance between the predicted and actual position of the curves is smaller when least-squares is used due to its smoothing effect. The distance also becomes smaller as the surface normal approaches the axis of rotation, as stated by Corollary 3. (d) Two images of a rotating toy. (e) The measured and predicted curves for three of the curves detected in the first frame. (f) Results of the classification process. Horizontal bars denote the occluding contour curves. The cross is on the side of the surface. The large diamond corresponds to the origin of the affine coordinate frame. The three small diamonds correspond to the points defining the three affine basis vectors.

in general, persist for large viewing direction changes; however, if at least four tangency points are visible during the observer's motion (not necessarily the same ones throughout), one can use an approach similar to that of Tomasi and Kanade [19] to handle occlusions. To improve tracking performance and reduce errors in the computation of $\|\mathbf{q} - \tilde{\mathbf{q}}\|$ we are also implementing a system for tracking curves rather than points.

8 Concluding Remarks

We are currently studying how this strategy can be used to derive more robust estimates of surface curvature. Cipolla and Blake [3] have shown that curvature estimates based on *rate of parallax* measurements are insensitive to observer acceleration and rotational velocity. In general, these estimates require surface markings to be identified in the vicinity of the occluding contour. We are studying an approach that eliminates the dependence on surface markings by computing the rate of parallax between the measured positions of image curves and the positions predicted by the stationarity assumption.

References

[1] Y. Aloimonos, "Active vision revisited," in *Active Perception*, pp. 1–18, Lawrence Erlbaum Associates, 1993.

[2] D. H. Ballard and C. M. Brown, "Principles of animate vision," in *Active Perception*, pp. 245–282, Lawrence Erlbaum

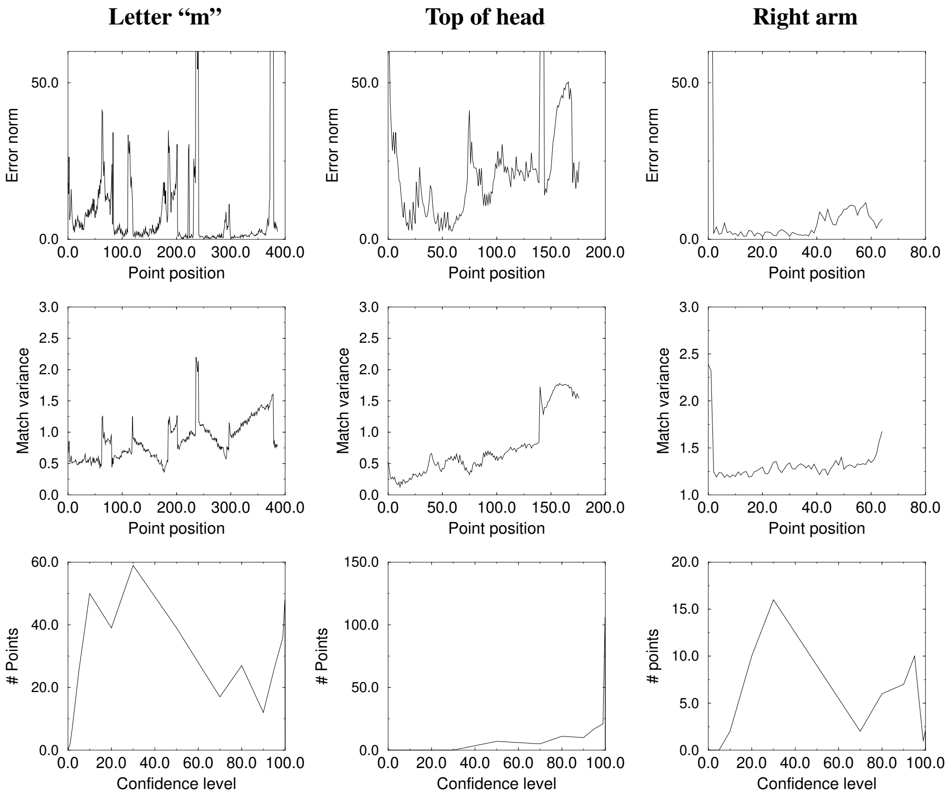

Figure 4: *First row:* Plot of the squared Mahalanobis distance between predictions and measurements for every point on the corresponding curves. *Second row:* The variance of the distance in image positions between matched points in consecutive frames. *Third row:* Histogram of the number of points on each curve having a given probability of being non-stationary.

Associates, 1993.

[3] R. Cipolla and A. Blake, “Surface shape from the deformation of apparent contours,” *IJCV*, 9(2): 83–112, 1992.

[4] J. J. Koenderink, *Solid Shape*. MIT Press, 1990.

[5] R. Vaillant and O. D. Faugeras, “Using extremal boundaries for 3-d object modeling,” *IEEE T-PAMI*, 14(2): 157–173, 1992.

[6] P. Giblin and R. Weiss, “Reconstruction of surfaces from profiles,” in *Proc. 1st ICCV*, pp. 136–144, 1987.

[7] J. J. Koenderink and A. J. van Doorn, “Affine structure from motion,” *JOSA-A*, no. 2, pp. 377–385, 1991.

[8] J. L. Mundy and A. Zisserman, eds., *Geometric Invariance in Computer Vision*. MIT Press, 1992.

[9] O. D. Faugeras, “What can be seen in three dimensions with an uncalibrated stereo rig?,” in *Proc. 2nd ECCV*, pp. 563–578, 1992.

[10] K. N. Kutulakos and C. R. Dyer, “Recovering shape by purposive viewpoint adjustment,” *IJCV*, 12(2): 113–136, 1994.

[11] A. Zisserman, A. Blake, C. A. Rothwell, L. J. van Gool, and M. van Diest, “Eliciting qualitative structure from image curve deformations,” in *Proc. 4th ICCV*, pp. 340–345, 1993.

[12] S. Ullman and R. Basri, “Recognition by linear combinations of models,” *IEEE T-PAMI*, 13(10): 992–1006, 1991.

[13] A. Shashua, “A geometric invariant for visual recognition and 3d reconstruction from two perspective/orthographic views,” in *Proc. IEEE Workshop on Qualitative Vision*, pp. 107–117, 1993.

[14] J. H. Rieger, “Three-dimensional motion from fixed points of a deforming profile curve,” *Optics Letters*, 11(3): 123–125, 1986.

[15] D. Weinshall and C. Tomasi, “Linear and incremental acquisition of invariant shape models from image sequences,” in *Proc. 4th ICCV*, pp. 675–682, 1993.

[16] K. N. Kutulakos and C. R. Dyer, “Occluding contour detection using affine invariants and purposive viewpoint control,” tech. rep., CS Department, University of Wisconsin—Madison, 1994.

[17] Z. Zhang and O. Faugeras, *3D Dynamic Scene Analysis*. Springer-Verlag, 1992.

[18] H. Cramér, *Mathematical Methods of Statistics*. Princeton University Press, 1946.

[19] C. Tomasi and T. Kanade, “Shape and motion from image streams under orthography: A factorization method,” *IJCV*, 9(2): 137–154, 1992.