MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo No. 1133 May, 1990

Feature Matching for Object Localization in the Presence of Uncertainty

Todd Anthony Cass

Abstract: We consider the problem of matching model and sensory data features in the presence of geometric uncertainty, for the purpose of object localization and identification. The problem is to construct sets of model feature and sensory data feature pairs that are geometrically consistent given that there is uncertainty in the geometry of the sensory data features. If there is no geometric uncertainty, polynomial-time algorithms are possible for feature matching, yet these approaches can fail when there is uncertainty in the geometry of data features. Existing matching and recognition techniques which account for the geometric uncertainty in features either cannot guarantee finding a correct solution, or can construct geometrically consistent sets of feature pairs yet have worst case exponential complexity in terms of the number of features. The major new contribution of this work is to demonstrate a polynomial-time algorithm for constructing sets of geometrically consistent feature pairs given uncertainty in the geometry of the data features. We show that under a certain model of geometric uncertainty the feature matching problem in the presence of uncertainty is of polynomial complexity. This has important theoretical implications by demonstrating an upper bound on the complexity of the matching problem, and by offering insight into the nature of the matching problem itself. These insights prove useful in the solution to the matching problem in higher dimensional cases as well, such as matching three-dimensional models to either two or three-dimensional sensory data. The approach is based on an analysis of the space of feasible transformation parameters. This paper outlines the mathematical basis for the method, and describes the implementation of an algorithm for the procedure. Experiments demonstrating the method are reported.

Acknowledgements: This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology, and was funded in part by an Office of Naval Research University Research Inititative grant under contract N00014-86-K-0685, and in part by the Advanced Research Projects Agency of the Department of Defense under Army contract number DACA76-85-C-0010, and under Office of Naval Research contract N00014-85-K-0124.

1 Introduction

A common approach to model-based object recognition is the hypothesis and verification paradigm. The pose in the image of a known object is hypothesized then evaluated and verified to localize the object in the image. The hypothesis stage often consists of deriving a set of geometric features from both the modeled object and the image data and determining correspondences between pairs of image and model features. From these correspondences a rigid transformation on the model features (and the model itself) is computed aligning the model geometrically with an hypothesized instance of the object in the image. This forms the object *pose* hypothesis which can then be evaluated and verified based on comparing the transformed model with the underlying image data.

The central problem in this method of hypothesis construction is determining the sets of corresponding image and model features. This feature matching is difficult in the domain of object localization for three main reasons. Due to object occlusions in the scene, some model features may have no corresponding image feature; some features are *missing*. Because there are other objects in the image, some image features will not correspond to any model features; some features are *spurious*. Finally, due to inaccuracies in sensing and feature extraction, there is some uncertainty in the geometry of image features; their measured positions or orientations may deviate from the correct values. These three factors, missing, spurious, and distorted image features conspire to make determining feature correspondences difficult.

In this paper we focus on the central problem of image and model feature matching in the presence of missing, spurious, and distorted features, for the construction of pose hypothesis. In particular we define a model of the geometrical uncertainty of image features, and devise a tractable algorithm for determining all *geometrically consistent* sets of feature correspondences given the uncertainty tolerances. The paper is organized into six main sections. Section 2 outlines the formal basis of the approach, and section 3 outlines details of the computation, and presents an algorithm for the construction of feature correspondences, and pose hypothesis. The final sections present some experimental results, extensions to this work, related work, and the conclusions of the paper.

2 The Idea: Approximate Matching

This paper considers the case of localization of 2D objects given 2D sensory data such as grey-level images. The problem at hand is to construct sets of feature correspondences from which to form pose hypotheses. We first define what constitutes a feature: the primary attributes of the features we use for pose hypothesis are a definite position and orientation in the plane. For the first part of this development we make use of point features, whose position characterizes the position of a point on the boundary contour of an object, and whose orientation may, for example, characterize the orientation of the contour normal at the point, if this is stable. We later extend the method to use line segments.

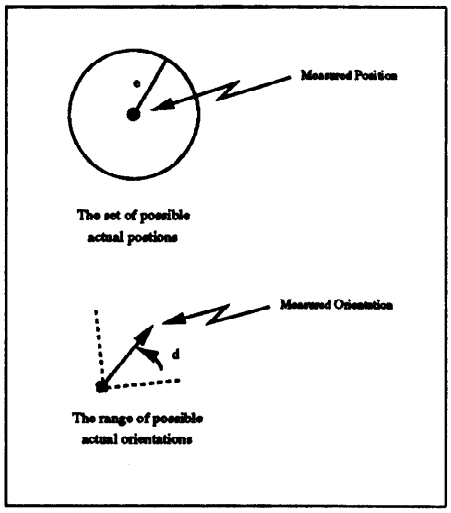

Figure 1: Shown are uncertainty bounds in position and orientation. Any position within a distance ϵ of the measured position is possible, as is any orientation within δ of the measured orientation.

2.1 The Basic Elements

This section presents the basic ideas involved in our approach to feature matching.

2.1.1 Bounded Uncertainty

A key assumption we make is that the uncertainty in the actual position and orientation of an image feature is bounded. This is a common assumption[8][4][3][7]. We consider two independent bounds on the positional and orientational uncertainty. Figure 1 illustrates these uncertainty bounds. The position of an image feature represents the measured position of the contour point. We assume the true position may deviate from the measured position by a maximum distance of ϵ, thus the real position falls within a circle of radius ϵ centered at the measured position. We assume the true orientation may deviate from the measured orientation by a maximum angle of δ, thus the real orientation falls within a range of orientations of length 2δ centered at the measured orientation.

2.1.2 Some Notation

We characterize a transformation by three parameters ϕ, u, and v, where ϕ is the angle of a rotation about the origin and u and v are translations in the x and y directions, respectively. In the 2D domain all positions and orientations can be represented by vectors $\vec{v} \in \Re^2$. Note that the vector space C of complex numbers is isomorphic to the vector space $\Re^2$. For much of the analysis it will be convenient to use complex numbers to represent positions and orientations. With this representation, rotation about the origin by an angle ϕ corresponds to multiplication by the complex exponential $e^{i\phi}$. If $\vec{a} = (a_x, a_y)^T$

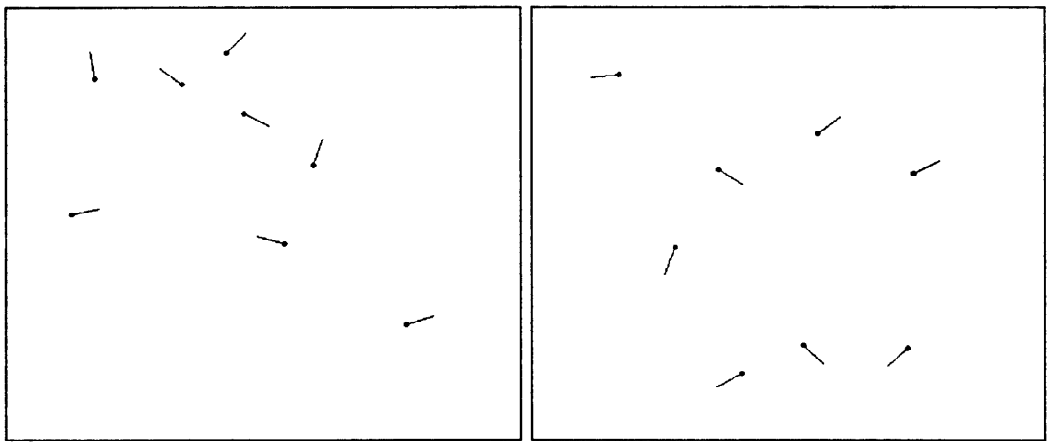

Figure 2: Model features on the left, image features on the right. A geometrically consistent match set is shown in figure 3. The positions are indicated by a small dot, and the orientations by a short line segment.

as a complex number we have $\vec{a} \leftrightarrow \mathbf{a} = a_x + ia_y$. Let the position of each model and image feature be represented by $\mathbf{p}_m$ and $\mathbf{p}_d$, respectively. Similarly, the orientation of the features are given by θ_m and θ_d. We will denote a feature match as an ordered pair (m,d), where $m = (\mathbf{p}_m, \theta_m)$ and $d = (\mathbf{p}_d, \theta_d)$. Finally, let $\mathbf{T}$ be the group of translations, $\mathbf{R}$ be the group of rotations about the origin, and the transformation group be $\mathbf{TPS} = \mathbf{T} \otimes \mathbf{R}$, their composition.[1] We will let $T_\phi \in \mathbf{R}$ be the operator for rotation by ϕ, and $T_t \in \mathbf{T}$ be the operator for translation by $\mathbf{t} = u + iv$. A transformation $T = T_t \circ T_\phi$ is the composition of a rotation and a translation, applied in that order.

2.2 Geometrically Consistent Match-Sets

We will use the term *feature match* or *match* to describe a single pair of a model feature and an image feature. An important fundamental definition is that of a *feasible transformation* for a particular feature match, which is closely linked to the notion of a *geometrically consistent* set of feature matches which we will introduce in this section. Let T be an arbitrary transformation, and let $(\mathbf{p}'_m, \theta'_m) = T[(\mathbf{p}_m, \theta_m)]$ represent a transformed model feature. For most $T \in \mathbf{TPS}$ we have $|\mathbf{p}'_m - \mathbf{p}_d| > \epsilon$, or $|\theta'_m - \theta_d| > \delta$. That is, after transformation the model and image features will not be even approximately aligned geometrically. But for some transformations the two matched features will be approximately aligned geometrically. Let $\mathbf{F}_{m,d} \subseteq \mathbf{TPS}$ be the set

$$\mathbf{F}_{m,d} = \{T \in \mathbf{TPS} : |\mathbf{p}'_m - \mathbf{p}_d| \leq \epsilon, |\theta'_m - \theta_d| \leq \delta, (\mathbf{p}'_m, \theta'_m) = T[(\mathbf{p}_m, \theta_m)]\},$$

that is, $\mathbf{F}_{m,d}$ is the set of transformations on the model feature m which leave it within ϵ in position, and δ in orientation of the data feature d. We define the set of *feasible transformations* for the feature match (m,d) to be the set $\mathbf{F}_{m,d}$.

[1] $\mathbf{TPS} = \mathbf{T} \otimes \mathbf{R}$ is the *semi-direct product* of the two subgroups.

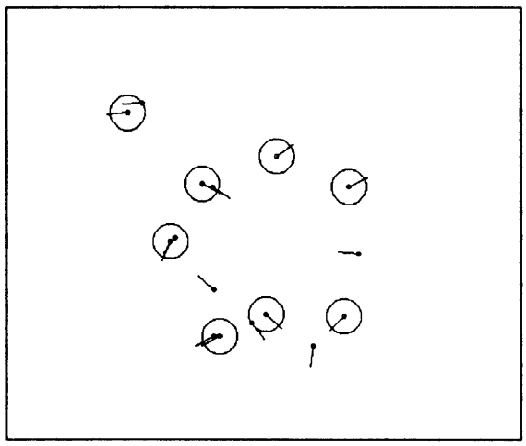

Figure 3: The model features transformed and plotted with the image features. The image features have uncertainty circles of radius ϵ around them. The model and image feature pairs which fall within ϵ and δ of one another form a geometrically consistent match set. The model and image features alone are shown in figure 2. In this picture, the model features from figure 2 have been rotated by about 170 degrees counterclockwise, translated and plotted over the image features.

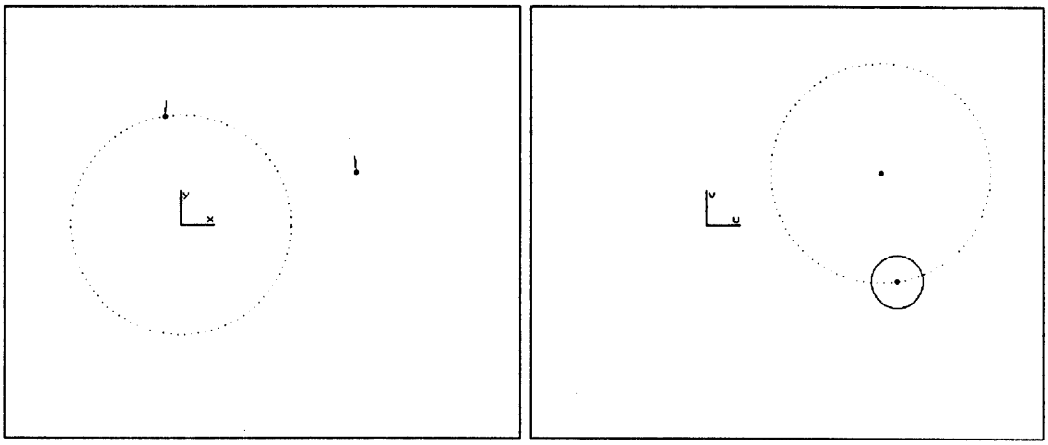

Figure 4: **Left:** the model and image features for a particular rotation, ϕ_0, of the model feature, shown in position space. The model feature's orbit under rotation is shown as a dotted circle. **Right:** the *match-disc* of feasible translations in translation space for this rotation of the model feature. Any translation $\mathbf{t} = u + iv$ within this disc will leave the rotated model feature within ϵ of the image feature. The dotted circle is the orbit of the match-disc as a function of the rotation of the model feature. Figure 7 shows this same rotation along with two nearby rotations and the effects on the match-disc.

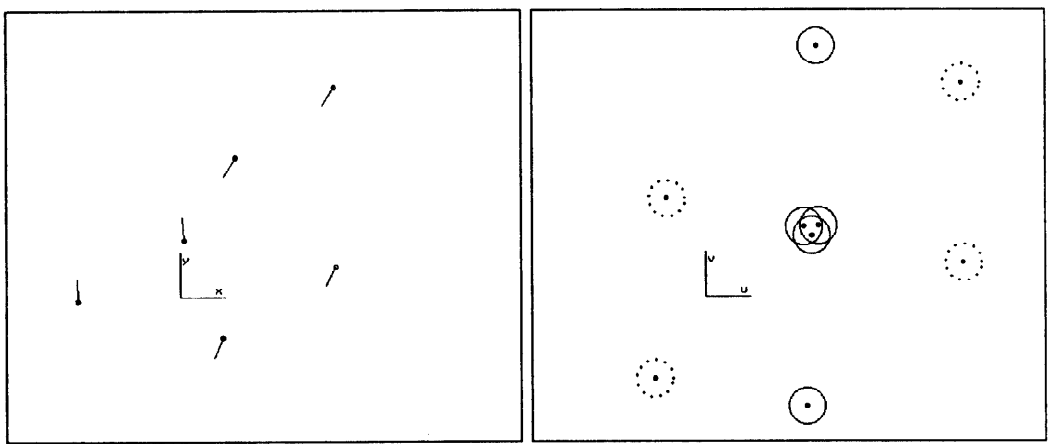

Figure 5: **Left:** Position space, showing the model features and the image features near the correct rotation of the model features; both are plotted together. **Right:** The match circles from all possible pairings of the model and image features. The dotted circles are not valid match-discs for the particular rotation because the difference in orientation for the feature match associated with these circles are are greater than δ. Note there are translations simultaneously aligning three matches.

We will use the term *match-set* to describe a set of model and image feature matches. Note that a given image or model feature can appear in several different matches in the same match-set, thus the mapping is not one-to-one. Let $M = \{(m_i, d_j)\}$ be a set of feature matches. Such a match-set is called *geometrically consistent* if

$$\bigcap_{(m_i,d_j)\in M} \mathbf{F}_{m_i,d_j} \neq \emptyset$$

that is if there exists some transformation which is geometrically consistent for all $(m_i, d_j) \in M$. Intuitively, the overlapping sets $\mathbf{F}_{m_i,d_j}$ for all $(m_i, d_j) \in \{m_i\} \times \{d_j\}$, divide **TPS** into equivalence classes where each class is associated with a different geometrically consistent match-set. More formally, let $T \in \mathbf{TPS}$ be a transform, and define $\varphi(T)$ to be the set $\{(m_i, d_j); T \in \mathbf{F}_{m_i,d_j}\}$[2]. The function $\varphi(T)$ partitions **TPS** forming equivalence classes E_k where $\mathbf{TPS} = \bigcup_k E_k$ and $T \equiv T' \iff \varphi(T) = \varphi(T')$. The set $\{E_k\}$ is the set of all maximal geometrically consistent match-sets. They're maximal in the sense that for each of them no other feature match can be added maintaining feasibility. Geometrically, $\mathbf{F}_{m,d}$ is isomorphic to a connected volume in $\Re^2 \times SO_2$ with a cylindrical tube shape. The union of the boundaries of these volumes divides $\Re^2 \times SO_2$ into cells, where each cell is isomorphic to an equivalence class E_k. Figure 6 illustrates several overlapping sets $\mathbf{F}_{m,d}$. Note that clustering techniques such as the Hough transform seek to find cells in **TPS** which are contained within many of the $\mathbf{F}_{m,d}$ by using a crude quantization of **TPS**.

The main goal then is to determine all maximal geometrically consistent match-sets. Each such match-set is associated with a set of geometrically consistent transformations,

[2]To be precise define $\mathbf{F}_0 = \mathbf{TPS} - \bigcup_{i,j} \mathbf{F}_{m_i,d_j}$. So for $T \in \mathbf{F}_0$, $\varphi(T) = \emptyset$

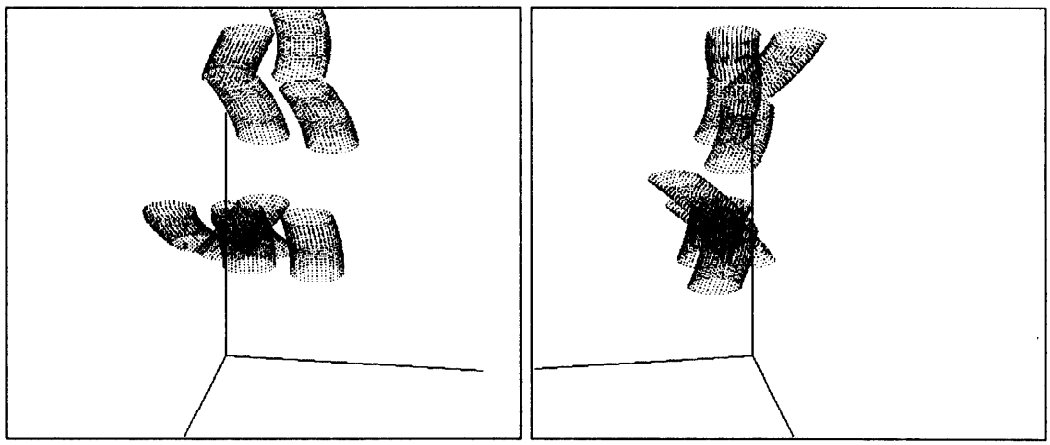

Figure 6: Two views of transformation parameter space. The vertical axis is rotation, ϕ, and the other two are translations u and v. Shown are a number of geometrically consistent transformation volumes for nine different feature pairs. Each set of consistent transformations is a helical tube, with extent in the vertical direction depending on δ the angle uncertainty, and circular cross section ϵ depending on the positional uncertainty. Slices of this space at fixed rotation ϕ are shown in figures 11 and 12.

and thus can be refined to form an hypothesis on the pose of the object in the image. We describe a method to enumerate the set $\{\varphi(T)|T \in \mathbf{TPS}\}$ of all maximal geometrically consistent match-sets. The algorithm is polynomial in the number of model and image features.

2.3 Topological Analysis of Transform Space

To introduce the feature matching and hypothesis technique we'll first consider the case of simple point features without any associated orientation. Thus the only relevant uncertainty bound is the uncertainty in image feature position ϵ. We'll then consider the case including orientation.

A crucial conceptual and algorithmic approach will be to consider the group of translations, $\mathbf{T}$, separately from the group of rotations $\mathbf{R}$. Define $\psi(T_t, T_\phi)$ to be the set $\{(m_i, d_j); T = T_t \circ T_\phi \in \mathbf{F}_{m_i,d_j}\}$. For a fixed rotation operator T_ϕ, the function $\psi(T_t, T_\phi)$ partitions the translation group $\mathbf{T}$ where as above, $T_t \equiv T_t' \iff \psi(T_t, T_\phi) = \psi(T_t', T_\phi)$. Let the equivalence classes be E_k^t, and define the function $\Psi(T_\phi) = \{E_k^t\}$ to be the set of equivalence classes of translations T_t for fixed rotation T_ϕ. The function $\Psi(T_\phi)$ partitions the set $\mathbf{R}$ of rotation operators, where $T_\phi \equiv T_\phi' \iff \Psi(T_\phi) = \Psi(T_\phi')$. Denote these equivalence classes by E_l^ϕ. This partition of $\mathbf{R}$ is crucial to the matching approach we develop here.

2.3.1 Feasible Translations

Consider an arbitrary but fixed angle of rotation ϕ_0, and a single feature match (m, d). If the model feature's position is rotated by an angle ϕ_0, the translation exactly aligning the two points is given by

$$\mathbf{t}_{ij} = \mathbf{p}_{d_j} - e^{i\phi_0}\mathbf{p}_{m_i}.$$

But we only require a transformation approximately aligning the two features. We call any translation T_t for which $|\mathbf{p}_{d_j} - T_t[\mathbf{p}_{m_i}]| \leq \epsilon$ a *feasible translation.* Define $\mathbf{C}_{ij}(T_\phi)$ be the set of feasible translations for feature match (m_i, d_j) after the rotation T_ϕ has been applied, thus

$$\mathbf{C}_{ij}(T_\phi) = \{\mathbf{t}_{ij} + \mathbf{z} : z \in \mathcal{C}, |z| \leq \epsilon, \mathbf{t}_{ij} = \mathbf{p}_{d_j} - T_\phi[\mathbf{p}_{m_i}]\}.$$

For each feature match, the set of feasible translations $\mathbf{C}_{ij}(T_\phi)$ for any fixed rotation T_ϕ is a disc in translation space; call this a *match-disc.* In this case, the regions of translation space formed when two or more discs have a common intersection are the equivalence classes of translation, E_k^t. Said another way, if we consider the fixed rotation ϕ_0 but consider all possible matches $\{(m_i, d_j)\} = \{m_i\} \times \{d_j\}$ and the set of associated match-discs $\{\mathbf{C}_{ij}(T_{\phi_0})\}$, the function $\psi(T_t, T_{\phi_0})$ partitions the translation space into equivalence classes E_k^t where $T_t \equiv T_t'$ when for all (i, j), $T_t \in \mathbf{C}_{ij}(T_{\phi_0}) \iff T_t' \in \mathbf{C}_{ij}(T_{\phi_0})$.

2.3.2 Topological Boundaries

We have been considering a fixed rotation ϕ_0. We now consider what happens to geometrically consistent match sets as the rotation ϕ varies. We showed that the center of each match-disc $\mathbf{C}_{ij}$ is the point in $\mathbf{T}$ given by $\mathbf{t}_{ij} = \mathbf{p}_{d_j} - e^{i\phi}\mathbf{p}_{m_i}$. Thus as ϕ varies, the match disc follows a circular path of radius $|\mathbf{p}_m|$ centered at the point $\mathbf{p}_d$ in the u-v plane. See figure 7. As the rotation ϕ varies the topology of intersections of match discs $\mathbf{C}_{ij}(T_\phi)$ in $\mathbf{T}$ changes; that is, the set $\Psi(T_\phi) = \{E_k^t\}$ changes. This partitions the rotation group $\mathbf{R}$. This is the key insight of the method. By partitioning $\mathbf{R}$, each equivalence class E_k^ϕ is associated with a partition of $\mathbf{T}$, together these form the partition of the whole transformation group $\mathbf{TPS}$ yielding the set of geometrically consistent match sets

$$\bigcup_T \{\varphi(T)\}.$$

Figures 8 and 9 show examples of different intersection topologies as the rotation ϕ is varied. Note that as yet we have not required a match set to be a one-to-one correspondence between model and image features. We will consider this below.

2.3.3 Determining Topological Boundaries

The function $\Psi(T_\phi) = \{E_k^t\}$ changes value when the topology of feasible match-disc intersections changes with variations in the rotation angle ϕ. There are three different events associated with this change. The first case is when two match-discs are intersecting and then move such that they are just tangent, and then completely separate. The ϕ at

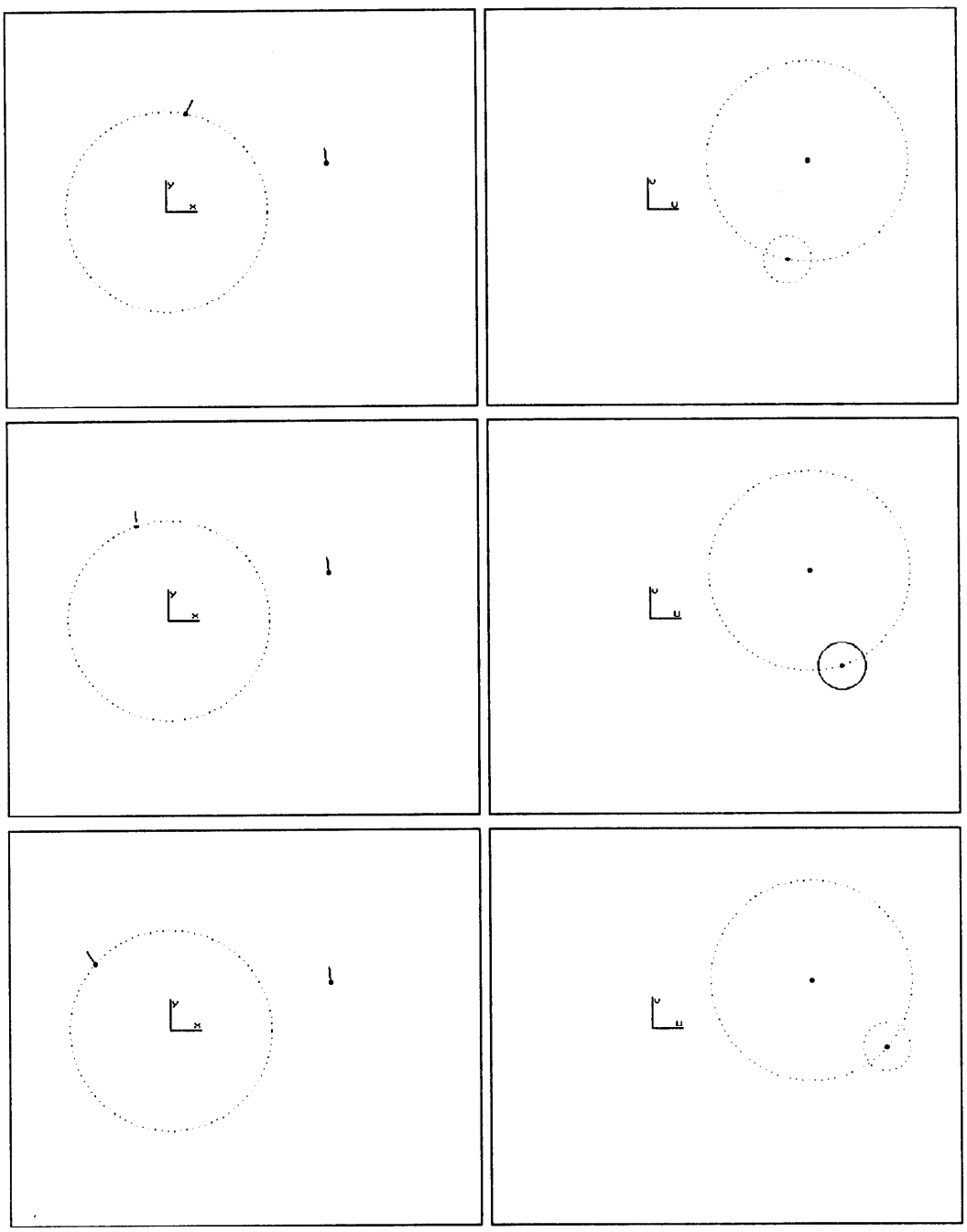

Figure 7: Snapshots of position space (the coordinates in which we represent feature positions) and translation space (the coordinates in which we represent relative translations) for three different rotations applied to the model feature. The match-circle in translation space is dotted when the difference in orientation between the image and model feature is greater than the uncertainty bound in orientation.

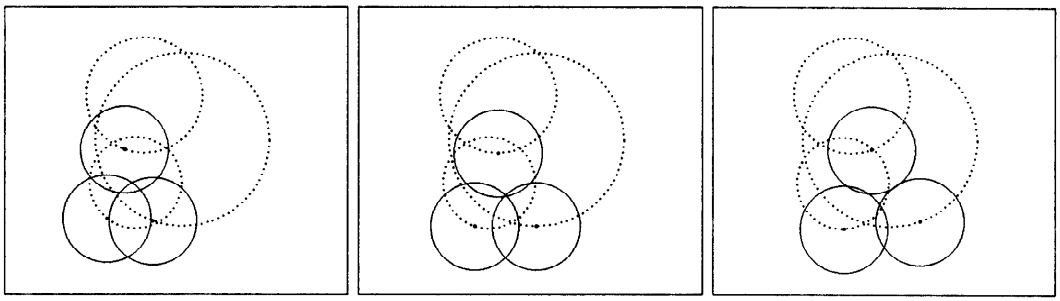

Figure 8: Three snapshots of translation space as match circles follow their circular orbits showing the case of a type I critical rotation due to a three-way intersection between three match circles. The solid circles are the match circles, the dotted circles are their orbits as ϕ varies. This illustrates how the topology of intersections changes as ϕ varies.

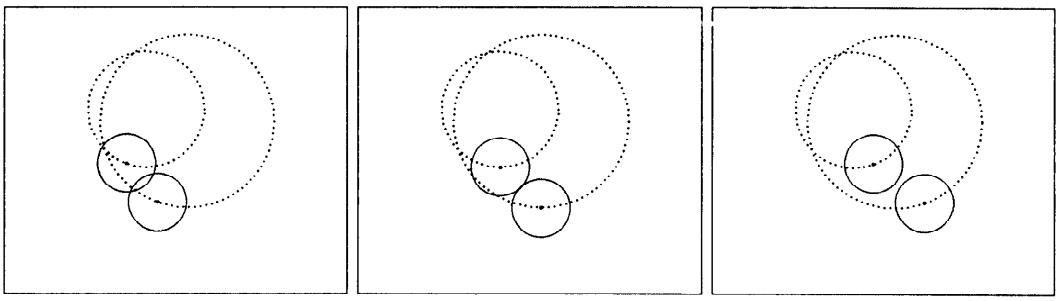

Figure 9: Three snapshots of translation space as match circles follow their circular orbits showing the case of a type I critical rotation due to tangency condition between two match circles. This illustrates how the topology of intersections changes as ϕ varies.

which tangency occurs marks a topological boundary. The second case is when three match-discs are mutually intersecting, then move such that they still intersect pairwise but have no common intersection. The ϕ at which the three boundary circles intersect at a single point marks a topological boundary. Finally, the ϕ's at which two match-discs exactly coincide marks a topological boundary. These three cases are illustrated in figures 8, 9 and 10. The fact that these are the only cases we need consider is discussed in the appendix in section 9. Considering all pairs and all triples of match-discs, we can compute those ϕ's marking topological boundaries in the rotation space, by analyzing all occurrences of the above three cases. Call these ϕ's *Type I critical rotations.* An example of a type I critical rotation is shown in figure 11. This figure is a slice of constant ϕ of the transformation space shown in figure 6.

When we include angle constraints, that is we require that $|\theta_d - T_\phi[\theta_m]| \leq \delta$, we introduce new topological boundaries. Figure 6 shows two views of **TPS** including the helical tubes forming geometrically consistent transformation sets for nine different feature

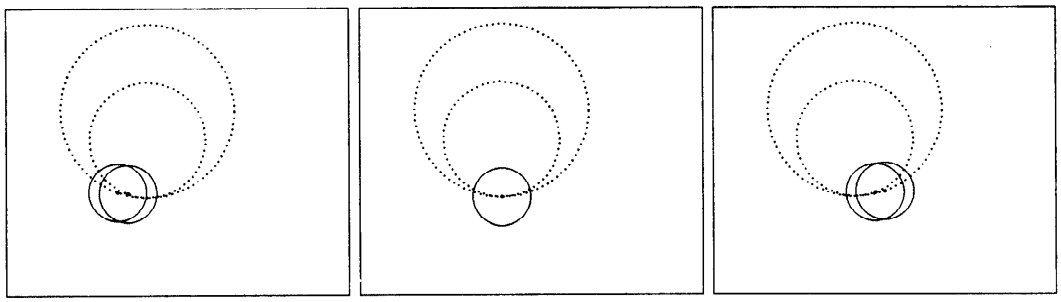

Figure 10: Three snapshots of translation space as match circles follow their circular orbits showing the case of a type I critical rotation due to a coincidence condition between two match circles.

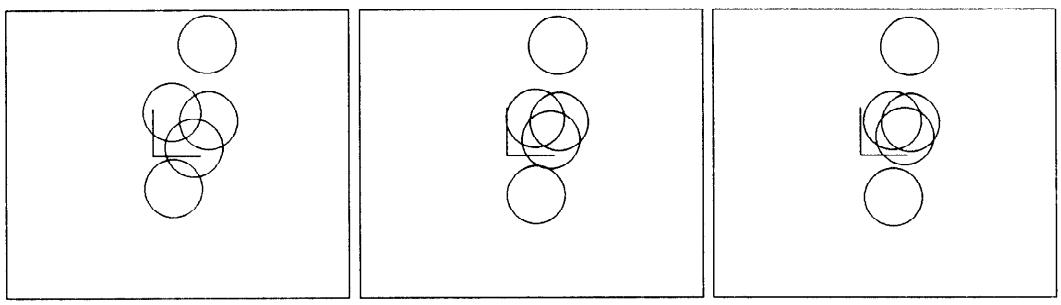

Figure 11: A sequence of match circles in translation space showing a type I critical rotation. These are slices of the same space as shown in figure 6. Type I critical rotations are when, as match circles orbit as functions of ϕ, their topology of intersection changes.

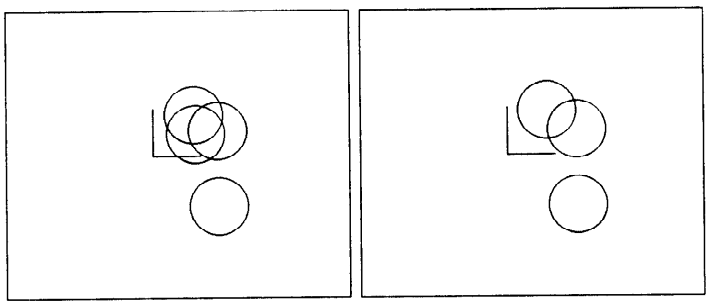

Figure 12: A sequence of rotations of match circles in translation space showing a type II critical point. These are slices of the same space as shown in figure 6. The topology of intersection of match-circles changes when a match-circle appears or disappears at the extrema of feasible orientation difference between the matched features.

matches. The ϕ's marking the ends of a given tube are also rotations where the topology of intersections of match-discs changes. Intuitively, a match-disc is present in **T** only in a rotation range in which the angle constraints are satisfied. The topology of **T**, and thus $\Psi(T_\phi)$, changes at the edge of the feasible rotation range for any feature match, where essentially a match-disc vanishes or appears. So for each feature match, the end points of the range feasible rotations also mark topological boundaries. Call these ϕ's *Type II critical rotations.* An example of a type II critical rotation is shown in figure 12.

2.3.4 Enumerating Match Sets

The critical rotations partition rotation space forming equivalence classes $\{E_k^\phi\}$ associated with constant topology of translation space. Each class $E_k^\phi = [\phi_{b,k}, \phi_{e,k}]$ is a range of rotation angles where $\phi_{b,k}$ and $\phi_{e,k}$ are critical rotations. To enumerate the match sets we use the fact that each class E_k^ϕ is associated with the set of translational equivalence classes $\Psi(T_\phi) = \{E_k^t\}$ for any $T_\phi \in E_k^\phi$. Consider one such equivalence class $[\phi_b, \phi_e]$. Pick any ϕ_m where $\phi_b < \phi_m < \phi_e$. At this rotation there is a particular configuration of match-discs in translation space. We'll call the circle forming the boundary of a match-disc a *match circle.* As we stated above, the union of the match circles divides the u-v plane into cells, each of which forms a translational equivalence class E_k^t. At a rotation ϕ_m as chosen above, if they intersect at all two circles intersect at exactly two points. By computing all these intersection points we have a set of points such that every cell in the u-v plane has at least two such points falling in its boundary. This gives us a point in each cell. Because $\phi_b < \phi_m < \phi_e$, at ϕ_m there are no 3-way match-circle intersections, circles just tangent to one-another, or coincident circles, and thus each match circle intersection point falls in the interior of all the match-circles it intersects except for the two match-circles from which it was computed, where it lies on the boundaries. The top and bottom frames of figure 11 illustrate generic configurations, while the middle frame shows a critical rotation delineating two equivalence classes.

We can enumerate all maximal geometrically consistent match-sets over all transformations, computing

$$\bigcup_{T_\phi, T_t} \{\psi(T_t, T_\phi)\},$$

as follows. For each equivalence class $[\phi_{b,i}, \phi_{e,i}]$ choose $\phi_{m,i}$ so $\phi_{b,i} < \phi_{m,i} < \phi_{e,i}$, apply the rotation $\phi_{m,i}$, and compute all match-circle intersection points. For each intersection point, determine the match-discs it falls within, building up a match set, and collect all such sets for all intersection points and all rotational equivalence classes. Albeit inefficient, this procedure can be used to enumerate the set of all maximal match sets.

2.4 Summary of the Idea

We have a set of model features and a set of image features, and the set of all possible pairs between them. We wish to find the subsets of pairs for which there exists some transformation on the model features simultaneously aligning them with their matched image feature, to within the uncertainty bounds on position and orientation. These we called geometrically consistent match sets. For any given rotation applied to the model features, and for each feature match, there is a disc in translation space of feasible translations. The intersections of these feasible match discs imply geometrically consistent match sets. We can partition the set of rotation angles ϕ into ranges in which the set of geometrically consistent match-sets does not change as ϕ varies within the range given by the partition. Thus there is a finite set of rotation angles ϕ which need be considered. For each such rotation, there is a finite set of geometrically consistent match-sets which we can compute from the individual feasible matches.

3 The Computation

The previous section outlined the basic idea of topological analysis of transform parameter space. This section describes the idea from a computational standpoint. We'll first describe the computational components needed, and then describe an algorithm for feature matching.

3.1 The Basic Components

3.1.1 Partitioning Rotation Space

We now give the details of the computation of critical rotations associated with change in the topology of feasible translations in translation space. There are three configurations of match-discs we need to consider: when two match-circles are just tangent, when three match circles intersect at a point yet have no common intersection of their interiors, and when two match-circles exactly coincide.

Two match-circles intersect at only one point when their centers $\mathbf{t}_1$ and $\mathbf{t}_2$ are exactly a distance 2ϵ apart. In this case the centers satisfy

$$(\mathbf{t}_1 - \mathbf{t}_2)(\mathbf{t}_1^* - \mathbf{t}_2^*) = 4\epsilon^2$$

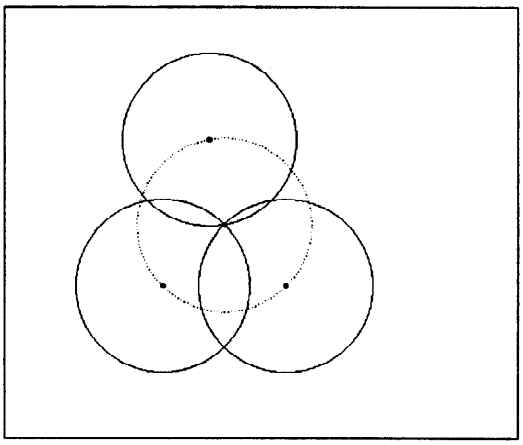

Figure 13: When three match circles intersect at a point, their centers fall on a particular circle of radius ϵ, centered at some point $\mathbf{t}_0$.

where $\mathbf{t}^*$ denotes the complex conjugate of $\mathbf{t}$. Thus we seek the roots of the equation

$$\mathbf{t}_1\mathbf{t}_1^* + \mathbf{t}_2\mathbf{t}_2^* - (\mathbf{t}_1^*\mathbf{t}_2 + \mathbf{t}_1\mathbf{t}_2^*) - 4\epsilon^2 = 0.$$

Recall that $\mathbf{t} = \mathbf{p}_d - e^{i\phi}\mathbf{p}_m$. Thus the above equation can be viewed as a function of $e^{i\phi}$. Multiplying the above equation by $e^{i\phi}$ will not change its roots, and inspection shows that the resulting equation is quadratic in $e^{i\phi}$ and therefore there are no more than two values of ϕ which satisfy it. If no unit magnitude complex number is a root then the circles do not intersect. Also, any ϕ will satisfy it when all the coefficients are zero. The appendix in section 9 gives the solution to the above equation for ϕ. There are mn circles so there at most $2\binom{mn}{2}$ such critical rotations.

Given three match circles whose centers are described by the complex quantities $\mathbf{t}_1(\phi)$, $\mathbf{t}_2(\phi)$, and $\mathbf{t}_3(\phi)$, we seek those ϕ for which they all intersect at a point. Note that this is equivalent to seeking those pairs $(\phi, \mathbf{t}_0)$ for which $\mathbf{t}_1$, $\mathbf{t}_2$, and $\mathbf{t}_3$ fall on a circle of radius ϵ centered at some point given by $\mathbf{t}_0$. See figure 13.

This case is described by the following system, which we must solve for ϕ and $\mathbf{t}_0 = u_0 + iv_0$:

$$\begin{aligned}
(\mathbf{t}_1(\phi) - \mathbf{t}_0)(\mathbf{t}_1^*(\phi) - \mathbf{t}_0^*) &= \epsilon^2 \\
(\mathbf{t}_2(\phi) - \mathbf{t}_0)(\mathbf{t}_2^*(\phi) - \mathbf{t}_0^*) &= \epsilon^2 \\
(\mathbf{t}_3(\phi) - \mathbf{t}_0)(\mathbf{t}_3^*(\phi) - \mathbf{t}_0^*) &= \epsilon^2
\end{aligned}$$

As is derived in the appendix in section 9, the values of ϕ which satisfy this system satisfy the equation

$$\begin{aligned}
&(\mathbf{t}_2 - \mathbf{t}_1)(\mathbf{t}_3 - \mathbf{t}_1)(\mathbf{t}_3^* - \mathbf{t}_2^*)(\mathbf{t}_2^* - \mathbf{t}_1^*)(\mathbf{t}_3^* - \mathbf{t}_1^*)(\mathbf{t}_3 - \mathbf{t}_2) + \\
&\qquad \epsilon^2[(\mathbf{t}_2^* - \mathbf{t}_1^*)(\mathbf{t}_3 - \mathbf{t}_2) - (\mathbf{t}_2 - \mathbf{t}_1)(\mathbf{t}_3^* - \mathbf{t}_2^*)]^2 = 0
\end{aligned}$$

Multiplying the above equation by $e^{3i\phi}$ will not change its roots, and inspection shows the result is a 6^{th} degree polynomial in $e^{i\phi}$. There are at most six distinct solutions to this, unless any ϕ is a solution when all the coefficients are zero. Note that since the above equation is purely real, the coefficients to powers of $e^{i\phi}$ and $e^{-i\phi}$, respectively, are complex conjugates; thus, the equation is a 3^{rd} degree trigonometric polynomial in $\cos(\phi)$ and $\sin(\phi)$. There are $\binom{mn}{3}$ triples of circles, so there are at most $6\binom{mn}{3}$ such critical rotations.

The third case where two match-circles exactly coincide is easily computed by equating $\mathbf{t}_1 = \mathbf{t}_2 = \mathbf{p}_{d_1} - e^{i\phi}\mathbf{p}_{m_1} = \mathbf{p}_{d_2} - e^{i\phi}\mathbf{p}_{m_2}$ and solving for ϕ. There are at most $\binom{mn}{2}$ such critical rotations. Finally, when we include constraints on relative orientation between matched features, there is a critical rotation where the difference in orientation of the model and image features is exactly δ, there are two such rotations for each feature match, thus exactly $2mn$ such critical rotations.

3.1.2 Constructing Match Sets

The critical rotations partition rotation space, and any equivalence class corresponds to a particular intersection topology of match-discs in translation space. As before, let an equivalence class be given by $[\phi_b, \phi_e]$ where these end points are adjacent critical rotations, and choose ϕ_m so that $\phi_b < \phi_m < \phi_e$. By the construction of the critical rotations, all the intersections of match circles at ϕ_m will be simply the two intersection points of pairs of match-circles. Given the locations of the centers of two match circles, $\mathbf{t}_1$ and $\mathbf{t}_2$, the two points of their intersection are given by

$$\frac{\mathbf{t}_1 + \mathbf{t}_2}{2} \pm i(\mathbf{t}_1 - \mathbf{t}_2)\beta$$

where

$$\beta = \sqrt{\frac{\epsilon^2}{(\mathbf{t}_1 - \mathbf{t}_2)(\mathbf{t}_1^* - \mathbf{t}_2^*)} - \frac{1}{4}}.$$

Each such point falls within some number of match-discs, and on the boundary of exactly two, those two that generated the intersection point. Thus each circle intersection point borders four different regions in translation space. See figure 14. To construct the different match sets associated with a given intersection point, we search through all the match-discs the point falls within, and then add either, neither, or both of the generating match circles to this set. Note that in enumerating the match-sets there is some duplication when two or more intersection points border the same cell in translation space.

We can now show a loose upper bound on the complexity of feature matching in this case. For each of the $O(m^3n^3)$ critical rotations there are at most $2m^2n^2$ intersection points. For each circle intersection point we must potentially look for containment in at most $mn - 2$ match circles in order to construct the match sets associated with it. So a crude upper bound on the complexity of constructing all maximal geometrically consistent match sets is $O(m^3n^3) \cdot 2m^2n^2 \cdot (mn - 2) = O(m^6n^6)$. This gives the basic idea, however there are more efficient ways to construct match sets which we will use in the algorithm development below.

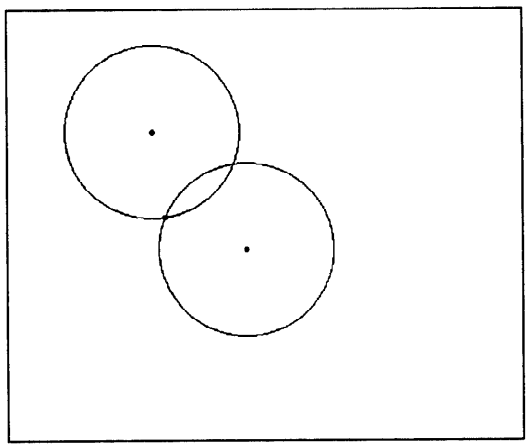

Figure 14: Each intersection point between two match circles borders four different regions of translation space, and is thus associated with four different match sets.

3.1.3 Evaluating Match Sets

Ultimately each match set forms a hypothesis on object pose which needs to be evaluated and verified in some way. As an initial step to this process we must determine the order in which the hypotheses will be verified, in case we don't verify them all. We seek to assign some value to each hypothesis. The simplest and most intuitive value is the number of matches in a match set. This value in a sense accounts for how well the hypothesis explains the observed image data. Note that as yet we have not required a match-set to be one-to-one. It is possible that different feature matches involving the same model or image features will be in the same geometrically consistent match set. In the case of point features it is likely that we require that features be matched one-to-one. Thus a straightforward count of the cardinality of a match set would be misleading.

One reasonable approach to this problem is to construct a one-to-one match set of maximal cardinality by eliminating some feature matches from the initial match sets. This is easy to do[9]. Construct one bipartite graph for each different match set. Let one set of nodes represent all the image features, and let the other set of nodes represent all the model features. There is an edge between two nodes if the given feature match is in the match set. The graph is bipartite because no edge connects two model feature nodes, or two image feature nodes. By finding a maximal bipartite matching we have a maximal one-to-one matching of matches from the match set. The bipartite matching problem takes $O(N^{2.5})$[10], where N is the number of nodes in the bipartite graph. Note that there may be many one-to-one match subsets of maximal cardinality, and we can only find one with the graph approach. However we only really seek a way to evaluate an hypothesis. Any match subset implies approximately the same range of transformations as its containing match set. The cardinality of a maximal one-to-one match set is the value assigned to each hypothesis.

3.1.4 Some Approximations

Thus far there is considerable complexity in the method as developed. First, to find all critical rotations takes time $O(m^3n^3)$ because we may have to consider all triples of match-discs. One reasonable approximation we can make is to only consider pairs of match discs instead. The idea is the following. We are looking for places where, say, N match circles have a non-empty intersection. By considering pairwise circle intersections, which we can find by computing type I critical rotations due to tangency between two circles, we can find a range of rotation where each of the N circles intersects the other $N-1$ circles pairwise. Although in this range we cannot be sure that the intersection of all N match discs is non-empty for these rotations, we will be very close to the rotation where this is true. Intuitively this is because the errors and thus the match-disc's positions are random variables and it is unlikely that the N discs will be randomly arranged so that they all intersect pairwise yet there is no region where all or most of them intersect. In section 4 we show that the empirical data support this hypothesis.

Second, when evaluating match-sets we need not perform a full bipartite matching. Instead we can form an upper bound on the size of the maximal match subset as follows. Determine the minimum of the number of distinct image features and the number of distinct model features appearing in a match set. This quantity is the upper bound we seek, eliminating all duplicate features. It can be computed in time linear in the number of matches in the match set. See [9] for a detailed discussion of this idea.

3.2 An Algorithm for Hypothesis Construction

This section details an algorithm for constructing all maximal geometrically consistent match-sets. We use the approximation that only pairs of match-disks are considered in determining critical rotations. In the actual implementation we make the following observation: Rather than choose an intermediate angle ϕ_m inside an equivalence class at which to evaluate *all* intersecting match circles we only look at regions of the u-v plane where change occurs. That is, we look at the regions containing the point of tangency of two match circles for Type I critical rotations and we look at all circle intersections contained with the match circle associated with a Type II critical rotation. The algorithm follows these basic steps:

- Form feature matches, and their corresponding match circles
- Extremes of valid rotation for each match circle form critical rotations
- Form pairs of match circles which intersect at some rotation
- Points of tangency for these pairs form critical rotations
- Compute and order these critical rotation angles
- Compute the match-sets geometrically consistent at some initial rotation angle, ϕ_0
- Step from ϕ_0 through critical rotation angles in order
- At each new critical rotation, determine any change in the match-sets formed

3.2.1 Intersecting Pairs of Match Circles

There are exactly mn feature matches, each associated with a match circle in translation space whose center is given as a function of rotation angle ϕ. If we consider the orbit of a match circle as a function of of ϕ without considering constraints on valid rotations due to orientation we have a circular orbit. Only over the range of rotations of width 2δ where the orientations of the model and image features are within δ of one another is the feature match feasible. See figure 7. Thus, most match circles will not intersect at any point in their orbit, and even fewer within their valid range of rotation angle ϕ.

To form pairs of intersecting match circles, the simplest approach is to consider all pairs of match-circles and determine if intersection is possible. There are $\binom{mn}{2}$ such pairs, requiring $O(m^2n^2)$ time. By being slightly more careful we can determine which circles could ever possibly intersect without necessarily considering all pairs of circles. We can bound the region of the u-v plane swept out by each match circle over its range of valid rotation by a rectangle, and determine which rectangles intersect. Only match circles whose bounding rectangles intersect can intersect one another. See figure 15. The intersections of N isothetic[3] rectangles in the plane can be computed in $O(N \lg N)+K$ time[12], where K is the number of intersecting rectangles. From these potential pairs the valid intersecting pairs

[3]Isothetic in this case means the sides of the rectangles are all aligned parallel to the coordinate axis

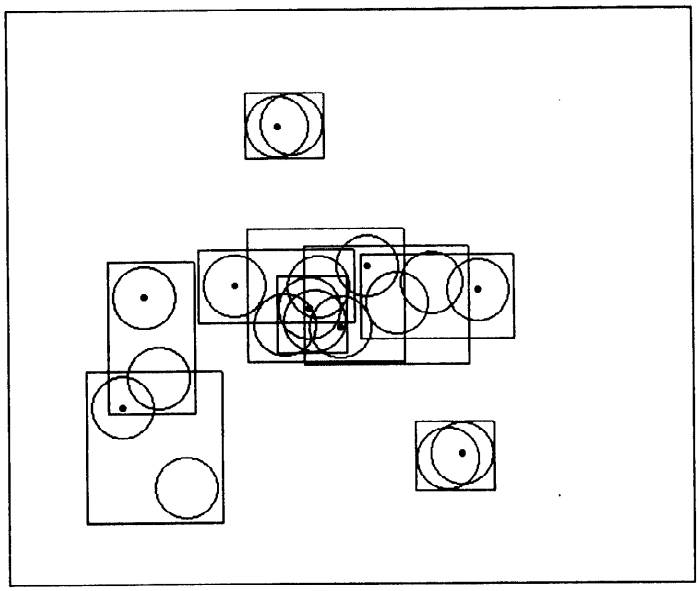

Figure 15: The rectangles bounding the area swept out by match circles over their valid range of possible rotations. Inside each rectangle are the positions of the match circle at each of its extremes in valid rotation. The position at the lower rotation angle is shown with a dot in the center of the circle.

can be determined. Of course, $K = O(N^2)$, but in practice there are many fewer rectangles which actually intersect, so this first step can reduce the computation considerably in practice.

3.2.2 Computing Maximal Geometrically Consistent Match Sets

For all match circles and all intersecting pairs of match circles, the set of critical rotations is computed and sorted modulo 2π. We pick an arbitrary rotation angle ϕ_0 that is interior to some partition, that is, ϕ_0 is not in the set of critical rotations. The idea will be to compute the geometrically consistent match sets at ϕ_0, and then step through the set of critical rotations noting any changes that occur in the set of geometrically consistent match sets. Changes in match sets can only occur at critical rotation points.

To start we compute the locations of the match circles at the rotation ϕ_0, and the points of intersections of match circles. As outlined in section 2.3.1, the union of the match circles divide the u-v plane into cells associated with match sets. Each distinct cell has at least two circle intersection points in its boundary. From these intersection points we can construct the match sets by counting the number of circles containing each intersection point. The appendix in section 9 describes an efficient algorithm for determining the set of match-sets for this initial, static configuration of match circles.

We construct and maintain a dynamic data structure containing, for each match-circle,

the set of other match-circles currently intersecting it. This table is updated at each critical rotation when the match-sets change. For a given circle producing an intersection point, these are the only circles that need be checked for containment of the intersection point in order to determine the associated match-set.

We next step through the equivalence classes of rotation space noting new geometrically consistent match sets as they appear. There are a few subtleties that must be considered. First consider two match-circles which intersect over some rotation range. Ignoring the angle constraints there are two type I critical rotations associated with these two match-circles. Also, each match circle also has a range $[\phi_{b_i}, \phi_{e_i}]$ over which it is valid. The intersection of these ranges $[\phi_{b_i}, \phi_{e_i}] \cap [\phi_{b_j}, \phi_{e_j}]$ for two different match circles yields the range of rotations where these circles are both valid. Note that either of the two type I critical rotations may or may not fall inside the range $[\phi_{b_i}, \phi_{e_i}] \cap [\phi_{b_j}, \phi_{e_j}]$. If a type I rotation, ϕ_c is not in $[\phi_{b_i}, \phi_{e_i}] \cap [\phi_{b_j}, \phi_{e_j}]$ then it is not a critical rotation we use to construct match sets, however we do use it to maintain the dynamic list of currently intersecting circles. Therefore, at each type I critical rotation we add or delete circles from the dynamic intersection sets as needed, but we also know whether or not any given match circle is within its range $[\phi_{b_i}, \phi_{e_i}]$ for the particular ϕ being considered. A second subtlety is that when a type II critical rotation occurs, several new match-sets are formed. All circle intersections falling *within* the new match-circle appearing must be considered to determine the new match-sets formed. Next we bring this all together in an algorithm.

3.2.3 Algorithm Summary

Considering all this we have the following procedure: Starting with the initial intersections noted at ϕ_0, step through the sorted critical rotations, updating the dynamic circle intersections lists at each Type I critical rotation. If a Type I critical rotation associated with circle i and circle j occurs within $[\phi_{b_i}, \phi_{e_i}] \cap [\phi_{b_j}, \phi_{e_j}]$ then compute the new match set implied by determining which circles contain the *single* point of intersection of the circles i and j for the critical rotation. Only the circles intersecting either i or j need to be considered. At each Type II critical rotation associated with a circle i, the pairwise intersection points of all circles intersecting circle i are computed, and for each a new match-set is constructed containing the match for circle i.

Finally, for each of the match sets constructed we can evaluate them to provide the output of the hypothesis module. Each match set was associated with some critical rotation, and one of the circle intersection points bounding the cell in translation space associated with the match sets provides a feasible translation. In this way with each match set we can associate a transformation hypothesis.

4 Experiments

We have shown that all maximal geometrically consistent match-sets can be found by analysis of transform parameter space, facilitated by partitioning rotation space at critical rotations. We have outlined and implemented the full algorithm which constructs match sets, as well as an approximate algorithm. The full algorithm computes all critical rotations;

the approximate algorithm does not consider the triple intersections described above but instead it only considers Type I critical rotations due to a tangency condition between two match circles in translation space, and all Type II critical rotations.

To explore the effectiveness of the approximate algorithm several empirical tests were performed. The question we need to answer is how often will a correct hypothesis be missed when the approximate algorithm is used. That is, given that there are k of the model features present in the image data, how often will one of the hypotheses include all k features, and what fraction of these k features will be found otherwise.

4.1 Simulations

In order to carefully control all aspects of the problem, simulation experiments were conducted in which synthetic model feature sets and data feature sets were constructed. Here we'll discuss how these data were constructed and the experiments run.

4.1.1 The Simulated Model and Image Features

Synthetic model and image features were constructed in the following way. In all the experiments point features were used consisting of a position in the plane and an associated orientation. For all experiments an "image" size of 256 by 256 was assumed; that is to say the coordinates of the position of a feature, (x, y), fell in the range $x, y \in [-128, 128]$. For any given experimental trial a total of m random points were generated with integer coordinates (x, y). To each of these points an independent unit vector of random orientation was assigned. This set of oriented feature points formed the model feature set. Examples of synthetic model and image features are shown in figure 2.

To construct simulated image data, a copy of the set of m model features was constructed. From this set, k of the m model features were chosen at random and the remaining features deleted from the set to simulate occlusions and missing features. For each remaining feature its position was perturbed by adding a random vector of length $l \leq (.95)\epsilon$, and its orientation was perturbed by a random angle chosen from $[-(.95)\delta, (.95)\delta]$. Next, an arbitrary rotation and translation were applied to this perturbed set. Finally, to simulate spurious data features additional random features were added to complete the synthetic image feature set.

4.1.2 The Experiments

Because the image and model features were constructed synthetically, it was possible to determine which matches were correct. Each experimental run determined whether one of the hypothesized geometrically consistent match sets contained all k correct feature matches among their matches, where k is the number of model features actually appearing among the image features. To demonstrate the effectiveness of the complete version of the algorithm in which the triple-circle intersection case was considered, the following experiment was run. $m = 15$ random model features were generated, from which $k = 11$ were randomly selected, independently perturbed in position and orientation, and added to 15 other random features. This set of $n = 26$ data features was then rotated and translated

by an arbitrary amount. This formed a model set of 15 features, and a data set of 26 features, of which 11 corresponded to the model. This process was repeated in 1000 independent trials using $\epsilon = 8$ and $\delta = 12$ degrees, generating completely new random model and data features each trial. When all critical rotations were computed, including triple intersections, a correct match-set of size 11 was found in all 1000 trials.

Because the existence of spurious features only adds to the number and size of match sets, but does not affect what fraction of the k model features appear in an hypothesized match set, for the following experiments there were no deleted or spurious features. The following experiments test the degree to which the *approximate algorithm* actually constructs all maximal geometrically consistent match sets.

4.1.3 The Effect of Model Feature Count

The first experiment explored the effect of increasing the number, m of model features. For these experiments there were $n = m$ image features. For each value of $m \in \{3, 6, 9, 12, 15, 18, 21, 24\}$, 10,000 *independent, completely random* experiments were performed. That is to say for each value of m, 10,000 different random model feature sets were generated, and from each of these the random image features were constructed by applying an independent random perturbation to each model feature, and then applying a rigid random rotation and translation to this set, independent from all other trials. The results of this experiment are summarized in two figures. Figure 16 shows how often the approximate algorithm failed to find a match-set with m correct feature matches in it, i.e. a correct hypothesis. Shown, for each value of m, is the fraction of 10,000 trials where a match-set containing m correct matches was not constructed.

Importantly, out of these 80,000 independent random experiments, a correct hypothesis was constructed in all but 7287 trials, of these, in 6970 trials the largest correct match-set was of size $m - 1$; in 313 trials the largest correct match-set was of size $m - 2$, and in only 4 trials was the largest correct match-set of size $m - 3$. This means that using the *approximate algorithm*, in 99.6% of the 80,000 trials a correct match-set of size m or $m - 1$ was constructed. Figure 17 shows the average over 10,000 trials of the fraction of the model correctly matched in the largest correct match-set, for each value of the model size m.

4.1.4 The Effect of Position Error and Uncertainty

The second experiment explored the effect of varying the amount of error in position which was introduced when constructing synthetic image features sets, and correspondingly varying the uncertainty bound assumed in the algorithm. For this experiment a range of integer position uncertainty bounds $\epsilon \in \{2, 4, 6, 8, 10, 12, 14, 16, 18, 20\}$ was used, and the error introduced in the synthetic image features was $(.95)\epsilon$. For each value of ϵ, 1000 independent trials were conducted, using 9 model features and 9 synthetic image features. Figure 18 shows how often the approximate algorithm failed to find a match-set with 9 correct feature matches in it, i.e. a correct hypothesis. Shown, for each value of ϵ, is the fraction of 1000 trials where a match-set containing 9 correct matches was not constructed. Out of 1000 trials, there were 851 trials where a correct match-set of size 9 was not con-

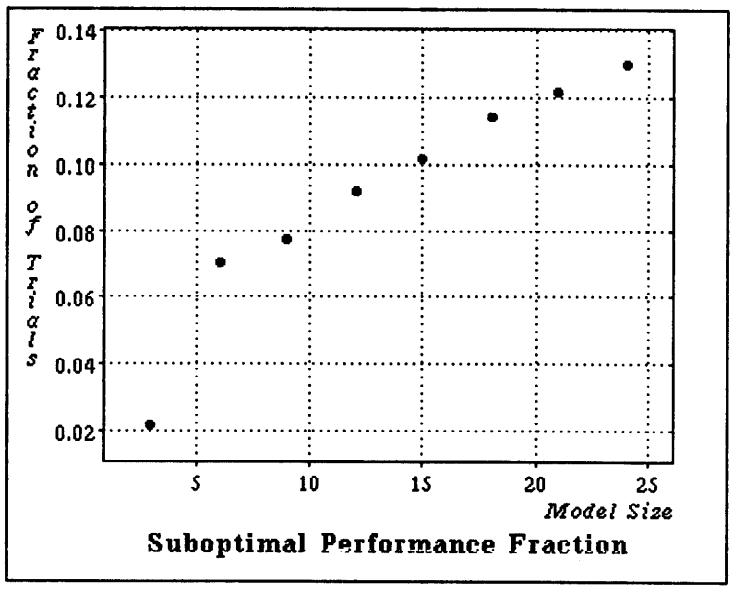

Figure 16: For each value of m, the number of model features, shown is the fraction of 10,000 trials where a match-set containing m correct matches was not found, although one or more existed.

structed, of which 828 constructed match-sets of size 8, and 23 constructed match-sets of size 7.

4.1.5 The Effect of Orientation Error and Uncertainty

The third experiment explored the effect of varying the amount of error in orientation which was introduced when constructing synthetic image features sets, and correspondingly varying the uncertainty bound assumed in the algorithm. For this experiment a range of integer orientation uncertainty bounds $\delta \in \{2, 4, 6, 8, 10, 12, 14, 16, 18, 20\}$ (in degrees) was used, and the error introduced in the synthetic image features was $(.95)\delta$. For each value of δ, 1000 independent trials were conducted, using 9 model features and 9 synthetic image features. Figure 19 shows how often the approximate algorithm failed to find a match-set with 9 correct feature matches in it, i.e. a correct hypothesis. Shown, for each value of δ, is the fraction of 1000 trials where a match-set containing 9 correct matches was not constructed. Out of 1000 trials, there were 735 trials where a correct match-set of size 9 was not constructed, of which 720 constructed match-sets of size 8, and 15 constructed match-sets of size 7.

4.2 A Real Example

Finally, as an example of a real application, and of running the algorithm on a much more complicated case, features were derived from real images. For both the model features and the image features, the features were derived by starting with a grey-level image of the

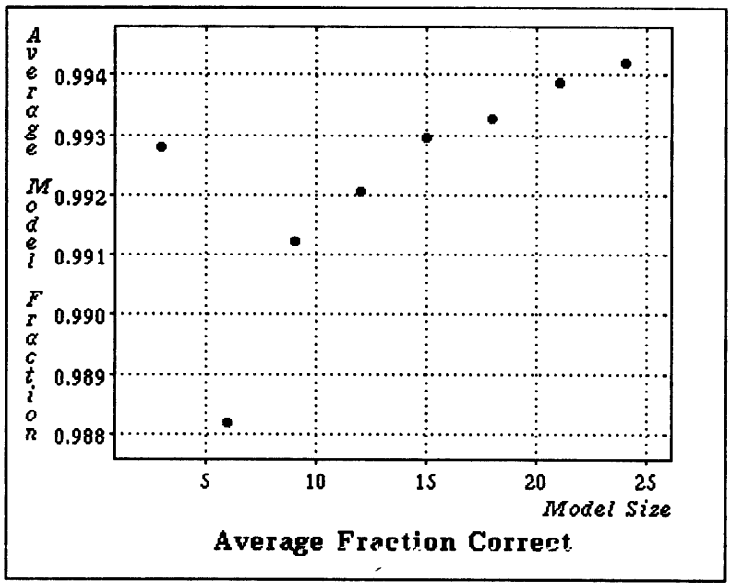

Figure 17: For each value of m, the number of model features, shown is the average over 10,000 trials, of the fraction of the model features correctly matched in the match set with the largest number of correct matches.

object and a scene, applying an edge detector, and sampling the resulting edge contours at regular intervals, forming features out of the sample point positions and the contour normal at that point. Of course, there are possibly much better ways to derive point features from edge contours, but this method demonstrates our point very well. Figure 20 shows the image from which the image data were derived. Figure 22 shows the edge contours for both the model and the input data image, and figure 21 shows the model and image point features derived from the edge contours. Figure 20 shows an example of one of the hypotheses constructed. The model contour is transformed according to the hypothesis and plotted over the image contours. In this case there were 41 model features and 161 image features, and the largest match-set had 27 matches. This is the match set we have displayed. This illustrates that we can use the method to derive robust pose hypotheses from complex images.

4.3 Discussion of the Experiments

The most important conclusion we can draw from these experiments is that the approximate algorithm, which only considers the interaction of pairs of match-circles, works nearly as well as the complete algorithm. Over a broad range of model size m, and the uncertainty parameters ϵ and δ, all image features were correctly matched to model features, except for one or two. When the model is reasonably large, say 10 features or more, almost all the model will be correctly matched.

The experimental results can be explained in terms of the interactions of match-circles.

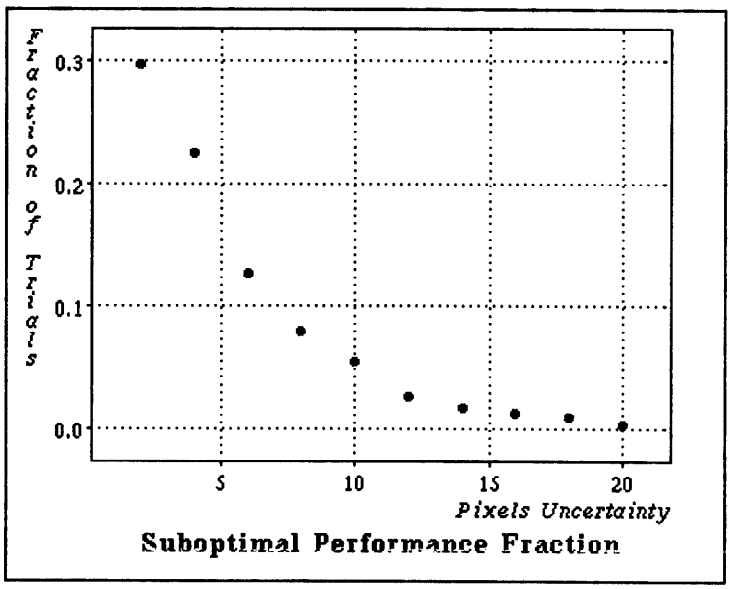

Figure 18: For the case of 9 model features and 9 image features, for each value of the position error introduced and the uncertainty assumed, we plot the fraction of 1,000 trials where a match-set containing 9 correct matches was not found, although one or more existed.

In the experimental results shown in figure 16, the fraction of cases with suboptimal performance rises with the model size m. This makes sense in retrospect. Imagine the u-v plane when the correct rotation has been applied. There is a region of intersection of all m *correct* match discs. As we vary ϕ and some disc moves off of this region of intersection, the more match-circles involved, the more likely it is to cross over the intersection of two other circles, rather than crossing only a single circle. In the case shown in figure 18, the larger the match-circles, the less likely many of them will actually bound the region of maximal intersection, rather than just containing it. Thus as some circle moves off there are fewer circle intersections for it to cross. Finally in figure 19 we see that the type II critical rotations reduce the need to look for type I critical rotations. The larger the angle uncertainty δ, the more circles interact with type I intersections, and thus the more likely we will need to consider a three-way intersection.

5 Extensions: Line Segments as Features

Many existing recognition approaches use straight line approximations to contours. One advantage of this approach is a great reduction in the number of features to process. The approach taken in this paper is not limited to point features, but applies equally well to features with finite extent such as line segments. First consider the case where the model is composed of straight line segments and the image data are oriented point features as before. This is closely related to the case where both the model and data are composed of

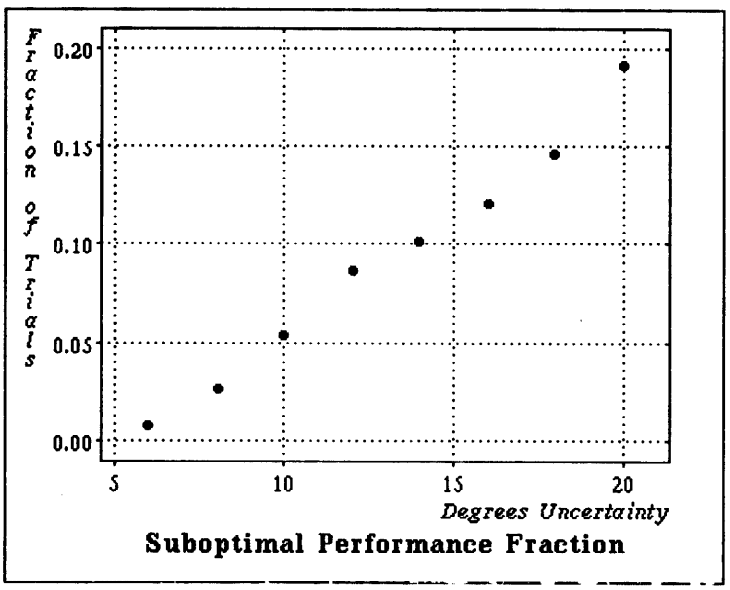

Figure 19: For the case of 9 model features and 9 image features, for each value of the orientation error introduced and the uncertainty assumed, we plot the fraction of 1,000 trials where a match-set containing 9 correct matches was not found, although one or more existed.

line segments if we subtract the length of the image segment from the model segment and treat the image segment as a point.

As before we will assume the uncertainty in position and orientation are independent. As in the point feature case, for a given feature match only those rotations leaving the orientation of the two features within δ of one another are feasible, where the orientation of a line segment is given by a unit perpendicular vector. In the case of position, consider the vector from the center of the line segment to the image point. See figure 23. For convenience we will define the range of feasible translations to be any translation leaving the perpendicular component (relative to the segment) of this vector less than ϵ, and the tangential component less than $\frac{l}{2} + \epsilon$ where l is the length of the model line segment. Thus the range of feasible translations, for any particular rotation, is a rectangular box.

Let $\mathbf{p}_m$ represent the midpoint of a model line segment. The center of a *match-rectangle* in translation space is given as before by

$$u + iv = \mathbf{t} = \mathbf{p}_d - e^{i\phi}\mathbf{p}_m$$

where $\mathbf{p}_d$ is the position of the image point. The match rectangle can be constructed by centering a rectangle around $\mathbf{t}$ whose sides are parallel and perpendicular to the orientation of the rotated model segment, $e^{i\phi}\theta_m$.

As we vary the rotation applied to a model segment, the rectangle of feasible translations rotates around the point $\mathbf{p}_d$ exactly as in the case described earlier. If we consider the translation space rectangles for all feature matches at any fixed rotation ϕ, the intersection topology of the rectangles defines geometrically consistent match sets as before. Again the

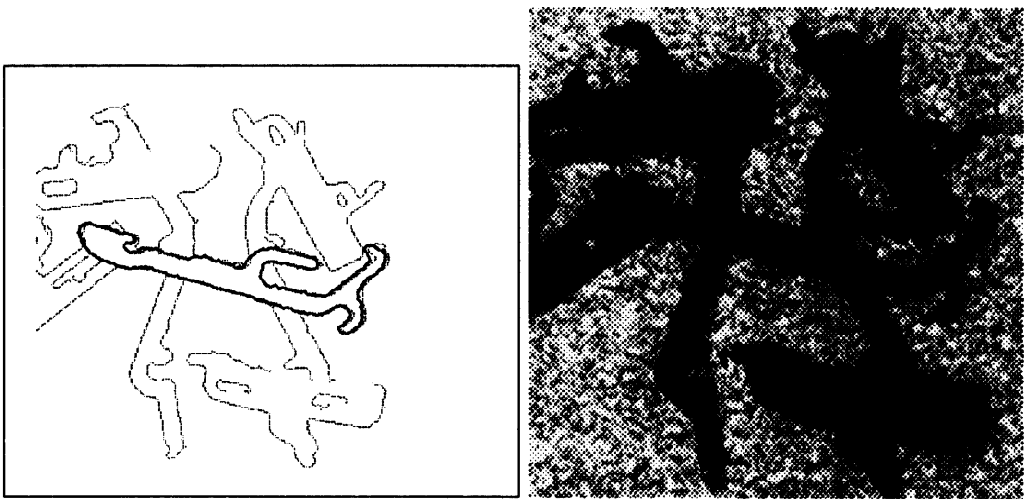

Figure 20: A real example. On the left is displayed the leading hypothesis. The model contour has been transformed and plotted over the image contours, indicating the pose hypothesized. On the right is the input image containing the object of interest.

idea is to partition the space of rotations into ranges within which the intersection topology of translation space is unchanged. As in the case with match-circles, the topology changes whenever two rectangles begin to intersect, and when three rectangles intersect at a single point, but have no common intersection. See figure 24.

Two rectangles begin to intersect when one of the vertices of one lies on a line segment of the other. We'll parameterize a line by a perpendicular unit vector $\mathbf{n}$, and perpendicular distance ρ from the origin to the line. The orientation $\mathbf{n}$ is given by $\mathbf{n} = -\sin\theta + i\cos\theta$, where θ is the positive angle from the x axis to the line. The equation for a line is then $\vec{x} \cdot \vec{n} = \rho$ in vector notation, or $\mathbf{x}\mathbf{n}^* + \mathbf{x}^*\mathbf{n} = 2\rho$ in our complex number representation. The equation $(\mathbf{x} - \mathbf{x_0})\mathbf{n_0}^* + (\mathbf{x}^* - \mathbf{x_0}^*)\mathbf{n} = 2\rho_0$ represents a line of the specified orientation whose perpendicular distance to the point $\mathbf{x_0}$ is ρ_0. The equation $(\mathbf{x} - \mathbf{x_0})e^{-i\phi}\mathbf{n_0}^* + (\mathbf{x}^* - \mathbf{x_0}^*)e^{i\phi}\mathbf{n_0} = 2\rho_0$ represents the same line rotated about $\mathbf{x_0}$ by an angle ϕ. In this way we can represent the position of a match-rectangle translation space by parameterizing its component line segments.

Two rectangles intersect in the manner of interest when a vertex of one rectangle lies in a line segment of another. Let the position of a vertex in translation space be given by

$$(\mathbf{v} - \mathbf{p}_{d_j})e^{i\phi} + \mathbf{p}_{d_j}$$

where $\mathbf{p}_{d_j}$ is the center of rotation of the match-rectangle in translation space, and $\mathbf{v}$ is the position in the u-v plane of a vertex when $\phi = 0$. Let the infinite line containing a line segment of a match-rectangle be given as above by

$$(\mathbf{x} - \mathbf{p}_{d_i})e^{-i\phi}\mathbf{n_i}^* + (\mathbf{x}^* - \mathbf{p}_{d_i}{}^*)e^{i\phi}\mathbf{n_i} = 2\rho_i$$

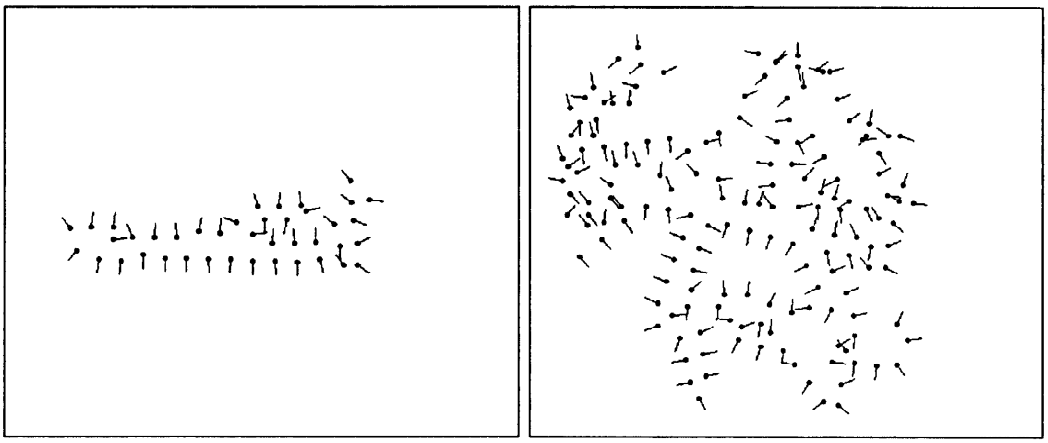

Figure 21: Model features on the left, image features on the right. These features were derived from real images as shown in figure 20

the point lies on the line when it satisfies the equation for the line so we substitute the first equation for $\mathbf{x}$ in the second equation. Let $\mathbf{v}_j = (\mathbf{v} - \mathbf{p}_{d_j})$. Substituting for $\mathbf{x}$ we get

$$(\mathbf{v}_j e^{i\phi} + \mathbf{p}_{d_j} - \mathbf{p}_{d_i})e^{-i\phi}\mathbf{n_i}^* + (\mathbf{v}_j^* e^{-i\phi} + \mathbf{p}_{d_j}{}^* - \mathbf{p}_{d_i}{}^*)e^{i\phi}\mathbf{n_i} = 2\rho_i$$

Note that all terms occur as complex conjugate pairs, so this equation is a real equation of the form $\alpha\cos\phi - \beta\sin\phi = \xi$. This can be solved by the equations $\phi = \pm\cos^{-1}(\frac{\xi}{\sqrt{\alpha^2+\beta^2}}) + \psi$ where $\psi = \tan^{-1}\frac{-\beta}{\alpha}$. Once we have solved for the ϕ where the vertex lies in the infinite line containing the line segment, we must check if it actually falls in the line segment. For a pair of match-rectangles, this check is done for all pairs of vertices from one rectangle to line segments of the other. We must also check that at the resulting ϕ the interior of the two rectangles have no common intersection, otherwise this is not a critical rotation in rotation space for these two rectangles.

In the case of a three-way rectangle intersection we solve for the intersection at a common point for three lines. Let the three equations be for $i = 1, 2, 3$

$$(\mathbf{x} - \mathbf{p}_{d_i})e^{-i\phi}\mathbf{n_i}^* + (\mathbf{x}^* - \mathbf{p}_{d_i}{}^*)e^{i\phi}\mathbf{n_i} = 2\rho_i.$$

We're solving for some $\mathbf{x}$ satisfying these equations for all 3 lines. Let $\mathbf{y} = \mathbf{x}e^{-i\phi}$ giving $\mathbf{y}\mathbf{n_i}^* - \mathbf{p}_{d_i}e^{-i\phi}\mathbf{n_i}^* + \mathbf{y}^*\mathbf{n_i} - \mathbf{p}_{d_i}{}^* e^{i\phi}\mathbf{n_i} = 2\rho_i$. The above system is three equations in three unknowns, two components of $\mathbf{x}$ and ϕ. Using gaussian elimination, eliminate $\mathbf{y}$ and $\mathbf{y}^*$ from the equations. Solve the resulting quadratic for $e^{i\phi}$, substitute back into the original equations and solve for $\mathbf{x}$. For each triple of match rectangles, all triples of line segments, one from each rectangle, must be considered. Finally, each intersection point solved for must lie inside each line segment, not just on its infinite extension; also at the resulting ϕ, the interiors of the rectangles must have no common intersection.

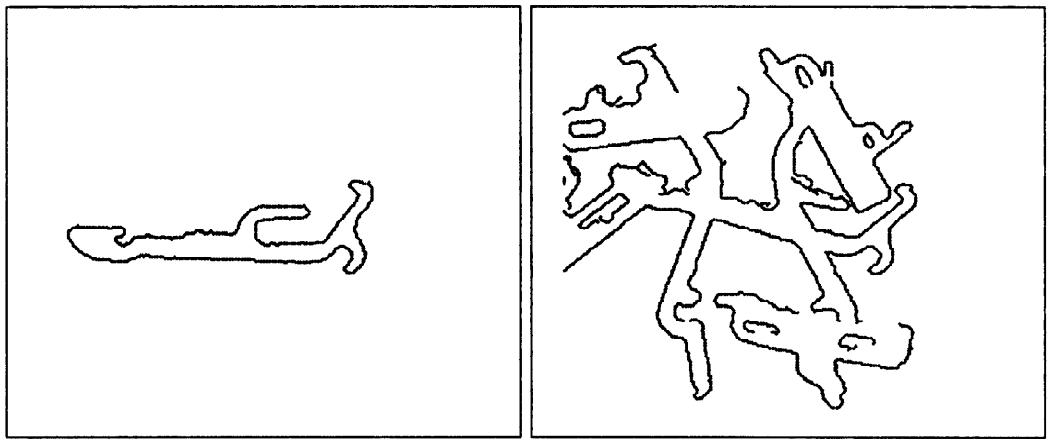

Figure 22: The edge contours derived from the model image and the input data image

These two cases serve to find all critical rotations in rotation space where topological changes occur in translation space. For each such critical rotation point, the intersection topology of rectangles in translation space must be examined to find the individual cells, each associated with a different match cell. These are found by finding the intersection points of all the line segments making up the rectangles. The rest of the procedure is analogous to the case of point features and match-circles.

It would be interesting to extend this idea to uncertainty cross sections of arbitrary convex polygons or curves. It may also be possible to have a multi-tiered uncertainty region, that is, cross sections that consist of several "concentric" convex closed curves, possibly weighted according to their likelihood. This would be a discrete approximation to a continuous uncertainty probability function.

6 Related Work

There has been a great deal of work on object identification and localization, particularly in the domain of 2D objects and sensory data. Some examples include [11][4][2][8][6]. The work most relevant to this paper is that which relies on determining the correspondences between local geometrical features derived from both the model and sensory data. Among these we consider here those approaches which explicitly account for error in the sensory data.

For the purposes of comparison let's state the important points about our approach. We utilize local geometric features characterized by a position and an orientation. We assume that there is uncertainty in the measured geometry of a feature derived from the sensory data. We further assume this uncertainty can be bounded. By analysis of the space of transformation parameters, we can construct all maximal geometrically consistent

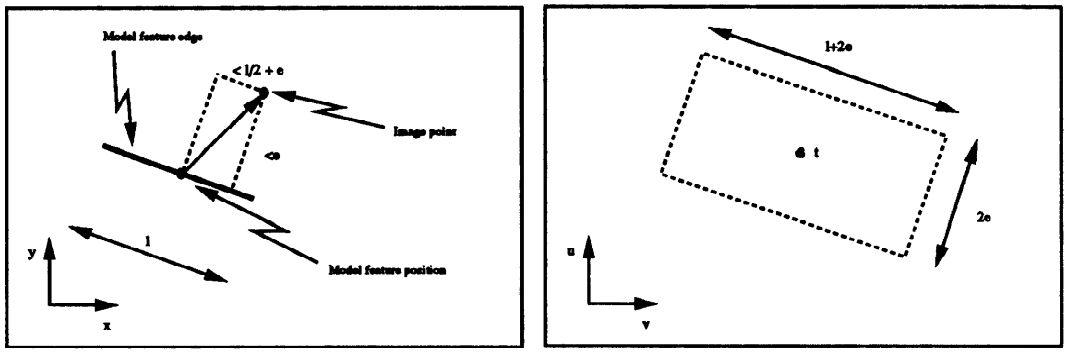

Figure 23: The position uncertainty for a line segment feature extends a distance ϵ in the direction perpendicular to the measured segment, and a distance $\frac{l}{2}+\epsilon$ from the measured center point in the direction parallel to the line segment. Above is shown the uncertainty in position space, below is shown the region of feasible translations in translation space.

match-sets in polynomial time[4]. Each match-set forms a globally consistent *interpretation* of the sensory data in terms of the model features. Note that without any approximation, these geometrically consistent match-sets can be found in time polynomial in the number of model and image features. The asymptotic runtime of the algorithm is independent of the geometry of any particular set of model or image features.

Grimson and Lozano-Perez[8] have developed a recognition system for both 2D and 3D objects they call RAF. The features they use are oriented line segments in the 2D case. They assumed independent bounds on positional and orientational uncertainty. In RAF the space of feasible sets of model and image feature pairs is explored sequentially, formulated as a tree search. Each path through the tree from root to leaf forms an element of the power set $2^{\{m_i\}\times\{d_j\}}$ of feature pairs, where each node in the tree corresponds to a particular feature pair. Large sections of the tree can be pruned away by considering, for each pair of feature matches, the intersection of their sets of geometrically consistent transformations. If any pair of matches is not a geometrically consistent match set, then the entire path rooted there can be ignored.

Empirically, the algorithm is quite robust, accurately finding objects in cluttered scenes given inaccurate sensory data. However because the algorithm is inherently exponential in the number of image features, several heuristics are used to make it tractable in practice. There are two main heuristics employed. One involves grouping feature matches into subsets via a Hough Transform form of parameter hashing and exploring the tree restricted to some of these subsets. The other involves an early cutoff of the tree search when a set of pairwise-consistent matches that is deemed good enough is encountered. The use of these two heuristics means that the entire space of match-sets is not explored. Finally, when a

[4]This assumes that the solutions to the sixth-degree polynomials is a constant time operation. However, as noted earlier, the approximate algorithm is very good, and the required quantities can be computed analytically.

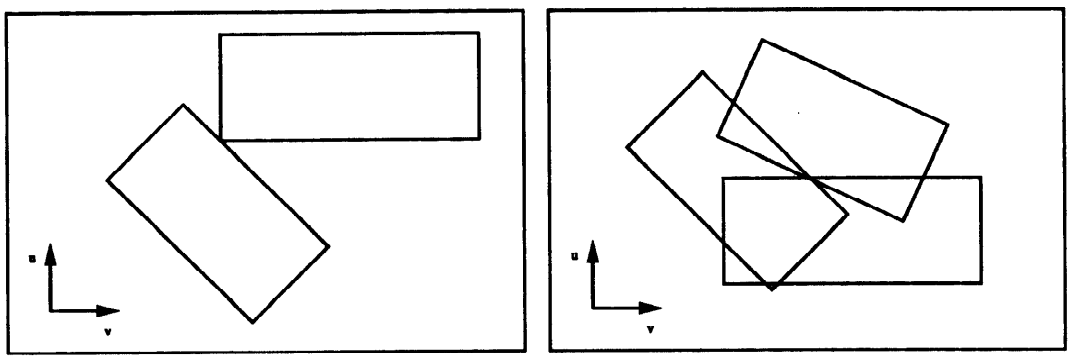

Figure 24: The events associated with topological change in translation space are two-way and three way intersections of rectangles as shown here.

pairwise geometrically consistent match-set is constructed an averaging method is used to derive a transformation from the set of feature matches. This is then applied to determine if in fact most feature matches are approximately aligned. Note that this does not determine if there exists a transformation simultaneously aligning all matches modulo uncertainty, i.e. that its a *geometrically consistent* match-set in our terminology. One difficulty with this search technique is that it is difficult to determine whether or not there are more instances of the object to be found after the first few are found. In fact if there are no instances it takes considerable search to answer negatively.

The principle advantages of our work over RAF are that it is worst case polynomial in the number of features, that it finds all maximal geometrically consistent match sets, and that these are by construction globally consistent. In the full case there is no approximation involved: all feasible matches within the uncertainty bounds are found. In particular, the hypothesis step hypothesizes all possible instances whether or not the object is present. In fact, if no object is present the computation is *easier* because in this case fewer match-circles will have common intersections. Note that one major difference in the two approaches is that RAF uses the more practical line segment features, while our system uses point features. In section 5 we outline how the approach can be generalized to line segment features, thus making the problems solved by the two systems largely equivalent.

Ellis[7] considers a special case of the method we discuss here. He assumes a model composed of line-segment features and data consisting of oriented point features. He assumes an uncertainty in position of magnitude ϵ, and an independent uncertainty in orientation of magnitude δ for the data features. *Given* a set of corresponding model and image feature pairs and uncertainty bounds he shows how to find the range of rotations on the data features which leaves them within ϵ of their paired model feature (or really within ϵ of the infinite line containing the line segment) and the difference in orientation of paired features within δ of one another. Then, given some rotation θ of the data features, he shows how to determine the range of translations on the data features allowed within the uncertainty in position.

The important part of this work is that he defines a model of uncertainty, and show how to localize an object with careful attention to the uncertainty. Ellis is considering a different problem where the correspondences are already known, which in a sense is a special case of the general matching problem. Using our approach to the analysis of transformation space we could accomplish the same task of localizing the object to within uncertainty bounds given the true correspondences. We would simply find the range of rotation ϕ within which all *correct* match-circles intersect in a common region.

In the field of computational geometry, Alt et.al.[1] describe a method for computing *approximate congruence* of two point sets of matched cardinality, allowing rigid transformations of the plane. This means given two points sets, A and B, of equal cardinality, find if there exists a one-to-one correspondence between them such that there exists a rigid transform of the points of set B bringing them within the ϵ neighborhood of their matched point from set A. They solve a very similar problem to ours. We might consider A to be image features, and B to be model features. They don't consider the case where A contains a only subset of B, as well as additional, unrelated points. They also determine only if at least one approximate congruence exists, instead of finding all approximate congruences of all sizes, as we do. They also do not consider angle uncertainty constraints as we do. It seems, however, that the modifications to their algorithm required to handle these cases is small. Thus the algorithm they outline could very likely be used to solve our matching problem in the case where angle constraints are not used. Their approach differs from ours in that they analyze the image space, instead of the transformation space. There are two possible advantages of our approach over theirs. First, it is clear how to extend our approach to extended features and polygonal uncertainty regions. Second, our method can be extended to higher dimensional transformations, such as including scale, while it is not clear how to extend their approach in these cases.

Baird[3] describes a method of matching features under uncertainty based upon linear programming. He assumes bounds on position uncertainty of image features, and outlines an expected polynomial time algorithm for matching. An important limitation of this work is that the size of the model and image feature sets is the same, that is, he does not allow missing or spurious features. He indicates that considering these cases substantially increases the complexity of the matching process.

The technique of *transformation sampling*[5] was the developmental ancestor to the analytic approach developed here. Rather than determine analytically all different sets of transformations, the space of transformations was sampled in hopes that a sample point would fall in each distinct feasible region of transformation space. The approach described here is also introduced briefly in [5].

The Hough transform clustering techniques[13] are really a crude approximation to our analytic approach. Again the idea is to determine where sets of feasible transformations intersect in transform space. But a coarse quantization is utilized to find such intersections, leading to considerable inaccuracy.

7 Conclusions

In summary we have shown a feature matching and pose hypothesis technique that requires time polynomial in the number of features, and is provably correct and complete for a certain class of local geometric features and precise models of geometric uncertainty.

We defined a *maximal geometrically consistent match set* as a set of model and image feature pairs for which there is some transformation which aligns the features in each match to within the bounds on geometric uncertainty, and such that at that transformation it is not possible to add more feature matches to the set and maintain feasibility. We showed that it is possible to compute all such geometrically consistent match-sets in time polynomial in the number of features.

One important implication is that a complete and precise solution to the problem of feature matching *in the presence of uncertainty* is of *polynomial* complexity. If the feature matching were exact, i.e. there where no error then there are simple polynomial algorithms to construct geometrically consistent match sets. However if only approximate matching is required then these simple algorithms cannot be proven to find all possible match-sets. Existing approaches which explicitly deal with geometric uncertainty as carefully as we do have been of worst case exponential complexity. Thus our analysis provides an important theoretical understanding of the matching problem.

This approach to matching is intended as a crucial part of the hypothesis stage of a recognition system. We have not addressed the fact that given a geometrically consistent match set, the actual transformation implied can be refined by optimizing some difference metric over the set of matches. Once high quality hypotheses are generated, the final step is to verify them using possibly richer representations.

While the 2D matching case described here is interesting, of greater interest is gaining a more thorough understanding of the matching problem with geometric uncertainty in higher dimensional problems such as matching 3D models to 2D image data, or 3D models to 3D range data. Our approach has important implications in these cases. Our analysis of transform space events goes beyond this simple case and can also be applied to these higher dimensional problems. An approach based on the ideas in this paper has been developed for the 2D case of rotation, translation including scale variation, and work is in progress considering matching 3D models to 2D and 3D data.

8 Acknowledgments

Thanks to Tao Alter, Eric Grimson, Berthold Horn, David Jacobs, and Tomas Lozano-Perez for reading earlier drafts and helpful comments and discussions.

9 Appendix

9.1 Solving three-way intersections

Given three match circles whose centers are described by the complex quantities $\mathbf{t}_1(\phi)$, $\mathbf{t}_2(\phi)$, and $\mathbf{t}_3(\phi)$, we seek those ϕ for which they all intersect at a point. Note that this is

equivalent to seeking those pairs (ϕ, t_0) for which t_1, t_2, and t_3 fall on a circle of radius ϵ centered at some point given by t_0. See figure 13. This case is described by the following system, which we must solve for ϕ and $t_0 = u_0 + iv_0$:

$$\begin{aligned}(t_1(\phi) - t_0)(t_1^*(\phi) - t_0^*) &= \epsilon^2 \\ (t_2(\phi) - t_0)(t_2^*(\phi) - t_0^*) &= \epsilon^2 \\ (t_3(\phi) - t_0)(t_3^*(\phi) - t_0^*) &= \epsilon^2\end{aligned}$$

By expanding then subtracting any two of the above equations we get

$$t_i t_i^* - t_j t_j^* - [t_0(t_i^* - t_j^*) + t_0^*(t_i - t_j)] = 0$$

From this we construct the linear system:

$$\begin{bmatrix} (t_2^* - t_1^*) & (t_2 - t_1) \\ (t_3^* - t_2^*) & (t_3 - t_2) \end{bmatrix} \begin{bmatrix} t_0 \\ t_0^* \end{bmatrix} = \begin{bmatrix} t_2 t_2^* - t_1 t_1^* \\ t_3 t_3^* - t_2 t_2^* \end{bmatrix}$$

The determinant of the above matrix is $(t_2^* - t_1^*)(t_3 - t_2) - (t_2 - t_1)(t_3^* - t_2^*)$ which is non-zero when t_1, t_2, and t_3 are not colinear, thus the solution of this system is the unique circle of some radius on which t_1, t_2, and t_3 fall. The case were the determinant is zero results when the three points lie on a circle of infinite radius. We can solve this for t_0 using Cramer's rule:

$$t_0 = \frac{\begin{vmatrix} (t_2 t_2^* - t_1 t_1^*) & (t_2 - t_1) \\ (t_3 t_3^* - t_2 t_2^*) & (t_3 - t_2) \end{vmatrix}}{\begin{vmatrix} (t_2^* - t_1^*) & (t_2 - t_1) \\ (t_3^* - t_2^*) & (t_3 - t_2) \end{vmatrix}}$$

which yields

$$t_0 = \frac{(t_2 t_2^* - t_1 t_1^*)(t_3 - t_2) - (t_3 t_3^* - t_2 t_2^*)(t_2 - t_1)}{(t_2^* - t_1 *)(t_3 - t_2) - (t_3^* - t_2^*)(t_2 - t_1)}$$

Note that the solution for t_0^* is consistent with the solution for t_0. We seek ϕ and t_0 such that this circle has radius ϵ. So, substitute t_0 into

$$(t_1(\phi) - t_0)(t_1^*(\phi) - t_0^*) = \epsilon^2$$

Now

$$(t_1 - t_0) = \frac{(t_2 - t_1)(t_3 - t_1)(t_3^* - t_2^*)}{(t_2^* - t_1^*)(t_3 - t_2) - (t_2 - t_1)(t_3^* - t_2^*)}$$

and finally

$$\begin{aligned}(t_2 - t_1)(t_3 - t_1)(t_3^* - t_2^*)(t_2^* - t_1^*)(t_3^* - t_1^*)(t_3 - t_2) + \\ \epsilon^2[(t_2^* - t_1^*)(t_3 - t_2) - (t_2 - t_1)(t_3^* - t_2^*)]^2 = 0\end{aligned}$$

9.2 Topological Changes in Translation Space

We claimed there were only three events which characterize the way in which the topology of intersection of match circles change as circles follow their circular orbits as functions of the rotation, ϕ, applied to the model features. The reason for this is fairly simple. Each distinct cell in translation space is bounded by one or more circles, and their intersection points. To change the topology, one cell must move into or out of another. To do so their boundaries must cross. When this happens either two, three, or more circles intersect at a single point. But if more than three intersect at a single point, any subset of three will do to solve the equations. There are also limiting cases where the cells of interest are just points, but these are handled as well. The case where two circle are coincident is the other special case which must be considered for topological effect.

9.3 Complex Numbers and 2D vectors

We find it convenient mathematically to consider points in the plane as complex numbers. With this representation, rotation about the origin by an angle ϕ corresponds to multiplication by the complex exponential $e^{i\phi}$. If $\vec{a} = (a_x, a_y)^T$ and $\vec{b} = (b_x, b_y)^T$, as complex numbers we have

$$\vec{a} \leftrightarrow \mathbf{a} = a_x + ia_y$$

$$\vec{b} \leftrightarrow \mathbf{b} = b_x + ib_y.$$

We also have

$$\vec{a} \cdot \vec{b} = \frac{\mathbf{ab^*} + \mathbf{a^*b}}{2}$$

$$\vec{a} \times \vec{b} = \frac{i(\mathbf{ab^*} - \mathbf{a^*b})}{2}$$

where $\mathbf{x^*}$ denotes the complex conjugate of $\mathbf{x}$.

9.4 Constructing The Initial Match-Sets

A first step in the algorithm is to determine the intersection topology of the match-circles for some initial rotation ϕ_0, and construct the associated match-sets. This can be done as follows. Compute the intersection points of all the circles, by first using an algorithm to intersect their bounding squares in $O(N \lg N) + K$ time for N circles with K intersection points. For each circle we sort its intersection points with other circles by angle. Because the circles are the same size, if the interior of two circles intersects, then the circles must intersect (excluding the special case of concentricity). Thus if we step through the intersection points in order of angle keeping track of when we enter or leave other circles, in two passes around each circle we can construct the set of match-sets associated with its intersection points.

References

[1] H. Alt, K. Mehlhorn, H. Wagener, and E. Welzl. Congruence, similarity, and symmetries of geometric objects. In *Discrete and Computational Geometry*, volume 3, pages 237–256. Springer-Verlag, New York, 1988.

[2] N. Ayache and O. D. Faugeras. Hyper: A new approach for the recognition and positioning of two-dimensional objecs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(1):44–54, 1986.

[3] Henry S. Baird. *Model-Based Image Matching Using Location.* MIT Press, Cambridge, MA, 1985.

[4] R.C. Bolles and R.A. Cain. Recognizing and locating partially visible objects: The local–feature–focus method. *International Journal of Robotics Research*, 1(3):57–82, 1982.

[5] Todd Anthony Cass. Parallel computation in model-based recognition. Master's thesis, Massachusetts Institute of Technology, 1988.

[6] Todd Anthony Cass. A robust implementation of 2d model-based recognition. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, Ann Arbor, Michigan, 1988.

[7] R. E. Ellis. Uncertainty estimates for polyhedral object recognition. In *Proceedings of IEEE Conference on Robotics and Automation*, pages 348–353, 1989.

[8] W. Eric L. Grimson and Tomas Lozano-Perez. Localizing overlapping parts by searching the interpretation tree. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(4):469–482, July 1987.

[9] Daniel P. Huttenlocher and Todd Anthony Cass. Measuring the quality of hypotheses in model-based recognition. *To Be Published*, 1990.

[10] Christos H. Papadimitriou and Kenneth Steiglitz. *Cominatorial Optimization: Algorithms and Complexity.* Prentice–Hall, Englewood Cliffs, New Jersey, 1982.

[11] W. A. Perkins. A model-based vision system for industrial parts. *IEEE Transactions on Computers*, 27(2):126–143, 1978.

[12] E. P. Preparata and M. Shamos. *Computational Geometry–An Introduction.* Springer-Verlag, New York, 1985.

[13] G. Stockman, S. Kopstein, and S. Bennet. Matching images to models for registration and object detection via clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 4(3), May 1982.