February 1991 LIDS-P-2020

COMPUTATIONAL LIMITATIONS OF MODEL BASED RECOGNITION*

Haim Shvaytser (Schweitzer)
Univ. of Texas at Dallas
P.O. Box 830688
Richardson, Texas 75083
haim@utdallas.edu

Sanjeev R. Kulkarni
MIT
Cambridge, MA 02139
kulkarni@lids.mit.edu, and
MIT Lincoln Laboratory
Lexington, MA 02173

Abstract

Reliable object recognition is an essential part of most visual systems. Model based approaches to object recognition use a database (a library) of modeled objects; for a given set of sensed data the problem of model based recognition is to identify and locate the objects from the library that are present in the data. We show that the complexity of model based recognition depends very heavily on the number of object models in the library even if each object is modeled by a small number of discrete features. Specifically, deciding whether a discrete set of sensed data can be interpreted as transformed object models from a given library is NP-complete if the transformation is any combination of translation, rotation, scaling, and perspective projection. This suggests that efficient algorithms for model based recognition must use additional structure in order to avoid the inherent computational difficulties.

*This work was supported by the U.S. Army Research Office under Contract DAAL03-86-K-0171 and by the Office of Naval Research under Air Force Contract F19628-90-C-0002.

1 Introduction

Many tasks of perceptual information processing that are easy and natural for humans appear to be much harder for machines. For example, although locating an object such as a pen on a table appears to us an easy task, it requires the ability to identify all possible shapes of pens as such, and is difficult to implement in a machine. These difficulties can be avoided in many computer vision applications that take place in a controlled environment. In these cases it is assumed that the objects of interest can be modeled and catalogued in a library. The problem of model based recognition can be informally described in the following way: given a library of modeled objects and a set of sensed data, identify and locate the objects from the library that are present in the data.

Reviews of the extensive literature on model based recognition in computer vision can be found in [1, 2, 3]; more recent studies include [5, 6, 9, 11]. The standard computational approach is to represent the modeled objects and the data in terms of discrete features so that the recognition can be solved as a search problem. These results indicate that by applying rigidity constraints in various ways, model based recognition can be efficiently applied to recognize a small number of object even from partial views and in the presence of non-malicious noise. The relevant complexity parameter in such cases is the number of features that model each object.

In this paper we analyze the case in which objects are represented by a small number of features. The relevant complexity parameter in this case is the number of objects. Instead of analyzing the performance of specific algorithms, our approach is to apply techniques from complexity theory to identify cases in which model based recognition appears to be inherently difficult. Specifically, we show that the problem is NP-complete, and thus, its complexity (modulo standard complexity assumptions, i.e., P $\neq$ NP) is exponential in the size of the library.

Proving that a problem is NP-complete is a common technique in complexity analysis for identifying the problem as intrinsically difficult. In a (well defined) sense, an NP-complete problem is the most difficult problem in the class NP, which includes many difficult problems such as the traveling salesman. However, an NP-complete problem is not completely unapproachable; a standard method for coping with such problems is to identify easily solved sub-problems. In the case of model based recognition this might correspond to exploiting additional structure of the modeled objects and the way they are viewed. For more information on the theory of NP-complete problems see [4]. For applications of NP-completeness results to vision tasks see [8, 10].

The negative results of this paper can be used to determine constraints that may simplify the problem of model based recognition. We will attempt to identify three types of constraints: constraints that leave the problem NP-complete, constraints that guarantee efficient (polynomial) algorithms, and constraints that make our NP-completeness proofs inapplicable, so that they may simplify the problem. The generic model based recognition problem that we consider is noise free and assumes no occlusion. An example of constraints of the first type is that every pair of local features can be found in at most three objects from the library. An example of constraints of the second type is that every pair of local features can be found in at most two objects from the library. An example of constraints of the third type is occlusion of convex objects.

2 Preliminary definitions

We consider situations in which objects can be described in terms of sets of local features. A local feature is a simple[1] geometric shape, and an object is described by a set of local features and their location in space. Commonly used features are points, lines, angles, etc. An example is shown in Figure 1, where a triangle is described in terms of straight lines (a), corners (b), and points along its edges (c).

Definition: An object description by local features is a set of pairs

$$O = \{\langle f_1, X_1\rangle, \langle f_2, X_2\rangle, \dots \langle f_t, X_t\rangle\}$$

where for $1 \leq i \leq t$, f_i is a local feature and X_i is its location is space relative to a fixed coordinate system.

Definition: A library is a set of object descriptions.

[1]The results of this paper hold for arbitrary interpretations of "simple" and "local".

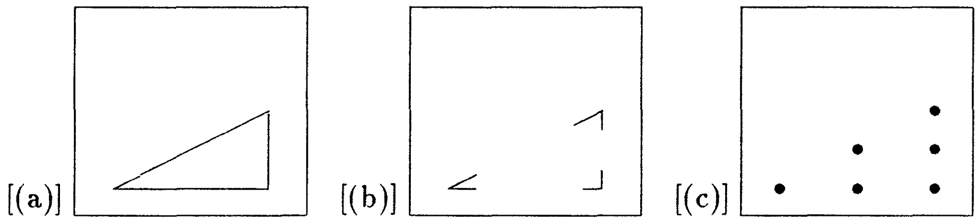

Figure 1: Examples of local features.

Definition: A picture is sensed data given as a set of local features and their location is space.

The problem of model based object recognition is:

> For a family of coordinate transformations Ψ, a library L, and a picture $P = \{\langle f_1, X_1\rangle, \ldots, \langle f_m, X_m\rangle\}$, determine a disjoint partition of P into objects from L, i.e., subsets $O_1, \ldots, O_q$ such that: (i) for $i \neq j$ $O_i \cap O_j = \emptyset$; (ii) $P = O_1 \cup \cdots \cup O_q$; (iii) for $1 \leq i \leq q$ there is $\psi_i \in \Psi$ that transforms an object from L into O_i.

Our main result is that the problem of model based recognition under translations, rotations, and perspective projections is NP-complete. The proofs are based on a reduction from *exact cover by 3 sets (X3C)* that is known to be NP-complete. (See [4] page 221.)

X3C: The following *exact cover by 3-sets* problem is NP-complete:

Instance: a set E of m elements and a collection C of 3-element subsets of E.

Question: does C contain an exact cover for E, i.e., a subcollection $C' \subset C$ such that every element of E occurs in exactly one member of C'?

Comment: X3C remains NP-complete even if no element occurs in more than three subsets in C, but is solvable in polynomial time if no element occurs in more than two subsets. A related problem, *exact cover by 2-sets*, is solvable in polynomial time.

$$p_1 \leftarrow m^2+1 \rightarrow p_2 \quad \cdots \quad p_i \leftarrow m^2+i \rightarrow p_{i+1} \quad \cdots \quad p_m$$

Figure 2: The picture in the proof of Theorem 1.

$$p_2 \leftarrow 2m^2+5 \rightarrow p_4 \leftarrow m^2+4 \rightarrow p_5$$

Figure 3: A typical object in the proof of Theorem 1.

3 The case of translation and rotation

Theorem 1: Let L be a library of objects and let P be a picture. The decision problem of whether P can be described as a disjoint union of translated and rotated objects from L is NP-complete. The problem remains NP-complete even if each object is described by 3 points.

Proof: Membership in NP is obvious. To show that the problem is NP-complete we reduce X3C to it.

Let $\{E, C\}$ be an instance of the X3C problem. C is a collection of 3-element subsets of the m elements $e_1, \ldots, e_m \in E$. We begin by constructing a picture P of m points $p_1, \ldots, p_m$ on the x axis. The location of p_1 is at the origin, the point p_2 is at distance m^2+1 from p_1, the point p_3 is at distance m^2+2 from p_2, etc. See the illustration in Figure 2. Let $\phi : E \rightarrow P$ denote the mapping of elements in E to points in P. For $1 \leq i \leq m$ we have:

$$\phi(e_i) = \text{a point at } x = (i-1)m^2 + i(i-1)/2 \tag{1}$$

Clearly, ϕ is 1-1 and onto, so that the inverse mapping is well defined. We now create the library L from the 3-element subsets in C. For a 3-set composed of the elements e_α, e_β, e_γ we add to L an object described by the 3 points $\phi(e_\alpha)$, $\phi(e_\beta)$, $\phi(e_\gamma)$. The object generated by the elements e_2, e_4, e_5 is shown in Figure 3.

To prove the NP-complete result it remains to show that P is a disjoint union of rotated and translated objects from L if and only if C contains an exact cover of E. The proof is based on Lemma 1 which is proved at the end of this section.

Let $C' \subset C$ be an exact cover of E, where $q = m/3 = |C'|$. For $\{e_{i_1}, e_{i_2}, e_{i_3}\} \in C'$ define $O_i = \{\phi(e_{i_1}), \phi(e_{i_2}), \phi(e_{i_3})\}$, so that $O_i \in L$ for $1 \leq i \leq q$. Since C' is a cover of E and ϕ is onto, $P = \bigcup_{i=1}^{q} O_i$. Since C' is exact and ϕ is 1-1, $O_i \cap O_j = \emptyset$ for $i \neq j$.

Conversely, let Ψ be the family of coordinate translations and rotations, and assume $O_i \in L$, $\psi_i \in \Psi$ for $1 \leq i \leq q$ such that: (i) for $i \neq j$ $\psi_i(O_i) \cap \psi_j(O_j) = \emptyset$; (ii) $P = \bigcup_{i=1}^{q} \psi_i(O_i)$. From Lemma 1 it follows that ψ_i is the identity transformation ($\psi_i(O_i) = O_i$), so that $\psi_i(O_i) \in L$ for $1 \leq i \leq q$. Let $O_i = \{p_{i_1}, p_{i_2}, p_{i_3}\}$. Define $T_i = (\phi^{-1}(p_{i_1}), \phi^{-1}(p_{i_2}), \phi^{-1}(p_{i_3}))$, and $C' = \{T_i : 1 \leq i \leq q\}$. From (ii) and the fact that ϕ^{-1} is onto it follows that C' is a cover. From (i) and the fact that ϕ^{-1} is 1-1 it follows that C' is an exact cover. $\square$

Lemma 1: Let O be an object from the library defined in the proof of Theorem 1, and let O' be an object defined by 3 points from the picture in the proof of Theorem 1. If O can be mapped by translation and rotation to O' then $O = O'$.

Proof: Without loss of generality let O be described by the points p_{i_1}, p_{i_2}, p_{i_3} and O' by the points p_{j_1}, p_{j_2}, p_{j_3}, where $i_1 < i_2 < i_3$ and $j_1 < j_2 < j_3$. Since the objects are 1-dimensional, a transformation taking O to O' involves either zero rotation or a 180° rotation. We show that the transformation must be with zero rotation and zero translation.

First, suppose the transformation involves no rotation, then the distance between p_{i_1} and p_{i_2} is the same as the distance between p_{j_1} and p_{j_2}. From Equation (1) we have

$$(j_2 - j_1)m^2 + \frac{j_2(j_2-1) - j_1(j_1-1)}{2} = (i_2 - i_1)m^2 + \frac{i_2(i_2-1) - i_1(i_1-1)}{2}$$

Let $s(i,j) = (j(j-1) - i(i-1))/2$, so that the above equation can be written as

$$[(j_2 - j_1) - (i_2 - i_1)]m^2 = s(i_1, i_2) - s(j_1, j_2). \qquad (2)$$

Clearly, $0 < s(i,j) < m^2$ for $1 \leq i < j \leq m$, and $|s(i_1,i_2) - s(j_1,j_2)| < m^2$. But since the right hand side of Equation (2) is divisible by m^2 it must equal

0, and we have

$$\begin{array}{c} s(i_1, i_2) = s(j_2, j_2) \\ j_2 - j_1 = i_2 - i_1 \end{array} \tag{3}$$

The unique solution to the system (3) with j_1, j_2 as the unknowns is $j_1 = i_1$ and $j_2 = i_2$. Since in pure translation the distance between p_{i_1} and p_{i_3} is the same as the the distance between p_{j_1} and p_{j_3} the same derivation gives $j_3 = i_3$, so that $O = O'$.

It remains to show that a transformation taking O to O' cannot involve rotation. Suppose, on the contrary, that O is mapped to O' by a transformation involving nonzero rotation. As mentioned above, this rotation must be 180°. But then the distance between p_{i_1} and p_{i_2} is the same as the distance between p_{j_3} and p_{j_2}, and the distance between p_{i_2} and p_{i_3} is the same as the distance between p_{j_2} and p_{j_1}. Using the same derivation as above we get $j_1 = i_3$, $j_2 = i_2$, and $j_3 = i_1$. But since $j_1 < j_3$ and $i_1 < i_3$ we have a contradiction. □

4 Translation, rotation, and scaling

Theorem 2: Let L be a library of objects and let P be a picture. The decision problem of whether P can be described as a disjoint union of translated, rotated, and scaled objects from L is NP-complete. The problem remains NP-complete even if each object is described by 6 points.

Proof: Membership in NP is obvious. To show that the problem is NP-complete we reduce X3C to it.

Let $\{E, C\}$ be an instance of the X3C problem. We begin by constructing a 2D picture Q as a disjoint union of two pictures: $Q = P \cup P'$. The pictures are defined by the two 1-1 and onto mappings: $\phi : E \rightarrow P$ and $\theta : E \rightarrow P'$.

$$\begin{array}{l} \phi(e_i) = \text{a point at } x = (i-1)m^2 + i(i-1)/2, \quad y = 0 \\ \theta(e_i) = \text{a point at } x = (i-1)m^2 + i(i-1)/2, \quad y = d \end{array} \tag{4}$$

See the illustration in Figure 4. We now create the library L from the 3-element subsets of C. For $(e_\alpha, e_\beta, e_\gamma)$ we add to L an object described by the 6 points: $\theta(e_\alpha)$, $\theta(e_\beta)$, $\theta(e_\gamma)$, $\phi(e_\alpha)$, $\phi(e_\beta)$, $\phi(e_\gamma)$. The object generated by the elements e_2, e_4, e_5 is shown in Figure 5.

$$
\begin{array}{ccccccccccccc}
p'_1 & & & & p'_2 & & p'_i & & & & p'_{i+1} & & p'_m \\
 & & & & \uparrow & & & & & & & & \\
 & & & & d & & & & & & & & \\
 & & & & \downarrow & & & & & & & & \\
p_1 & \leftarrow & m^2+1 & \rightarrow & p_2 & \cdots & p_i & \leftarrow & m^2+i & \rightarrow & p_{i+1} & \cdots & p_m
\end{array}
$$

Figure 4: The picture in the proof of Theorem 2.

$$
\begin{array}{ccccccccc}
p'_2 & & & & p'_4 & & & & p'_5 \\
\uparrow & & & & & & & & \\
d & & & & & & & & \\
\downarrow & & & & & & & & \\
p_2 & \leftarrow & 2m+5 & \rightarrow & p_4 & \leftarrow & m+4 & \rightarrow & p_5
\end{array}
$$

Figure 5: A typical object in the proof of Theorem 2.

To complete the proof it remains to show that Q is a disjoint union of translated, rotated, and scaled objects from L if and only if C contains an exact cover of E. The proof is based on Lemma 2 which will be proved at the end of this section.

Let $C' \subset C$ be an exact cover of E, with $q = |C'|$. For $\{e_{i_1}, e_{i_2}, e_{i_3}\} \in C'$ define $O_i = \{\theta(e_{i_1}), \theta(e_{i_2}), \theta(e_{i_3}), \phi(e_{i_1}), \phi(e_{i_2}), \phi(e_{i_3})\}$, so that $O_i \in L$ for $1 \leq i \leq q$. Since C' is a cover of E, and ϕ, θ are onto P and P' respectively, $Q = P \cup P' = \bigcup_{i=1}^{q} O_i$. Since C' is exact and ϕ, θ are 1-1, $O_i \cap O_j = \emptyset$ for $i \neq j$.

Conversely, let Ψ be the family of coordinate translations rotations and scaling and assume $O_i \in L$, $\psi_i \in \Psi$ for $1 \leq i \leq q$ such that: (i) for $i \neq j$, $\psi_i(O_i) \cap \psi_j(O_j) = \emptyset$; (ii) $Q = \bigcup_{i=1}^{q} \psi_i(O_i)$. From Lemma 2 it follows that ψ_i is the identity transformation, so that $\psi_i(O_i) \in L$ for $1 \leq i \leq q$. Let $O_i = \{p_{i_1}, p_{i_2}, p_{i_3}, p'_{i_1}, p'_{i_2}, p'_{i_3}\}$, where we assume without loss of generality that $p_{i_1}, p_{i_2}, p_{i_3}$ have zero y coordinates. Define $T_i = \{\phi^{-1}(p_{i_1}), \phi^{-1}(p_{i_2}), \phi^{-1}(p_{i_3})\}$, and $C' = \{T_i : 1 \leq i \leq q\}$. From (ii) and the fact that ϕ^{-1} is onto E it follows that C' is a cover. From (i) and the fact that ϕ^{-1} is 1-1 it follows that C' is an exact cover. $\square$

Lemma 2: Let O be an object from the library defined in the proof of Theorem 2, and let O' be an object defined by 6 points from the picture in the proof of Theorem 2. If O can be mapped by translation, rotation, and scaling to O' then $O = O'$.

Proof: Let O be generated by $e_{i_1}, e_{i_2}, e_{i_3}$. Let u_1, u_2, u_3, be the points of O' that are mapped to $\theta(e_{i_1}), \theta(e_{i_2}), \theta(e_{i_3})$ respectively, then u_1, u_2, u_3 are collinear. Similarly, let v_1, v_2, v_3, be the points of O' that are mapped to $\phi(e_{i_1}), \phi(e_{i_2}), \phi(e_{i_3})$ respectively, then v_1, v_2, v_3 are collinear. Since $\theta(e_{i_1})$, $\theta(e_{i_2})$, $\phi(e_{i_1})$ form a right triangle, u_1, u_2, v_1 form a right triangle, so that the triplets u_1, u_2, u_3, and v_1, v_2, v_3 are not on the same line in the picture. Therefore, it must be that one triplet lies on the line $y = 0$, and the other on the line $y = d$, and since the distance between the lines in the library object is d, the transformation involves no scaling.

It remains to show that the transformation involves no translation and rotation and this follows from Lemma 1 when applied to the points u_1, u_2, u_3 and the library of objects defined by the triplets of points $\{\theta(e_{i_1}), \theta(e_{i_2}), \theta(e_{i_3})\}$ for $1 \leq i \leq q$. $\square$

5 The case of perspective projection

A perspective projection is the mapping $\pi : \mathcal{R}^3 \to \mathcal{R}^2$ given by

$$x = f\frac{X}{Z} \quad ; \quad y = f\frac{Y}{Z} \tag{5}$$

Here it is assumed that the camera is at the origin and pointed directly down the Z axis. The reference frame is oriented as the image plane, which is located at distance f from the origin. See [7]. Unlike translation, rotation, and scaling, perspective projection may destroy geometric properties by merging lines and points. In the extreme case, any object far enough from the image plane is projected into a single point in a finite resolution picture. To eliminate degenerate cases we consider only *stable* perspective projections.

Definition: A stable perspective projection has the following properties: (i) Distinct 3D feature points are mapped into distinct 2D feature points. (ii) Non-collinear 3D feature points are mapped into non-collinear 2D feature points.

Notice that a small perturbation of the viewing point of an unstable perspective projection always gives a stable perspective projection.

Theorem 3: Let L be a library of 3D objects, and let P be a 2D picture given as a set of local features and their 2D location. The decision problem of whether P can be described as a stable perspective projection of a disjoint union of translated and rotated objects from L is NP-complete. The problem remains NP-complete even if each object is described by 12 points.

Proof: Membership in NP is obvious. To show that the problem is NP-complete we reduce X3C to it.

Let $\{E, C\}$ be an instance of the X3C problem. We begin by constructing the 2D picture $Q = P_1 \cup P_2 \cup P_3 \cup P_4$, where

$$P_j = \{\phi_j(e_i); 1 \leq i \leq m\} \text{ for } 1 \leq j \leq 4$$

$$\begin{aligned}
\phi_1(e_i) &= \text{a point at } x = (i-1)m^2 + i(i-1)/2, \quad y = 0 \\
\phi_2(e_i) &= \text{a point at } x = (i-1)m^2 + i(i-1)/2, \quad y = m^3 \\
\phi_3(e_i) &= \text{a point at } y = (i-1)m^2 + i(i-1)/2, \quad x = -1 \\
\phi_4(e_i) &= \text{a point at } y = (i-1)m^2 + i(i-1)/2, \quad x = m^3 + 1
\end{aligned}$$

Thus, the points are on the edges of a planar rectangle.

We now create the library L from the 3-element subsets of C. For $(e_{i_1}, e_{i_2}, e_{i_3})$ we add to L an object described by the twelve 3D points: $\{\phi_j^z(e_{i_t}); 1 \leq j \leq 4,\ 1 \leq t \leq 3\}$, where:

$$\begin{aligned}
\phi_1^z(e_i) &= \text{a point at } X = (i-1)m^2 + i(i-1)/2,\ Y = 0,\ \ Z = f \\
\phi_2^z(e_i) &= \text{a point at } X = (i-1)m^2 + i(i-1)/2,\ Y = m^3,\ Z = f \\
\phi_3^z(e_i) &= \text{a point at } Y = (i-1)m^2 + i(i-1)/2,\ X = -1,\ Z = f \\
\phi_4^z(e_i) &= \text{a point at } Y = (i-1)m^2 + i(i-1)/2,\ X = m^3 + 1,\ Z = f
\end{aligned}$$

Observe that $\pi(\phi_j^z(e_i)) = \phi_j(e_i)$ for $1 \leq j \leq 4$. It remains to show that Q is a stable perspective projection of a disjoint union of translated and rotated objects from L if and only if C contains an exact cover of E. The proof is based on Lemma 3 which will be proved at the end of this section.

Let $C' \subset C$ be an exact cover of E, with $q = |C'|$. For $\{e_{i_1}, e_{i_2}, e_{i_3}\} \in C'$ define O_i as the 3D object described by the twelve 3D points: $\{\phi_j^z(e_{i_t}); 1 \leq j \leq 4,\ 1 \leq t \leq 3\}$, so that $O_i \in L$ for $1 \leq i \leq q$. Since C' is a cover of E, and ϕ_j are onto P_j respectively, $Q = \bigcup_{i=1}^q \pi(O_i)$. Since C' is exact and ϕ_j are 1-1, $\pi(O_i) \cap \pi(O_j) = \emptyset$ for $i \neq j$.

Conversely, let Ψ be the family of coordinate translations and rotations and assume $O_i \in L$ and $\psi_i \in \Psi$ for $1 \leq i \leq q$, such that: (i) for $i \neq j$, $\pi(\psi_i(O_i)) \cap \pi(\psi_j(O_j)) = \emptyset$; (ii) $Q = \bigcup_{i=1}^q \pi(\psi_i(O_i))$. From (ii) and Lemma 3 it follows that ψ_i is the identity transformation, so that $\psi_i(O_i) \in L$ for $1 \leq i \leq q$. Let $O_i = \{p_{i_1}^j, p_{i_2}^j, p_{i_3}^j\}$ for $1 \leq j \leq 4$, where we assume without loss of generality that $p_{i_t}^j$ were generated by ϕ_j^z. Define $T_i = \{(\phi_j^z)^{-1}(p_{i_t}^j) : 1 \leq j \leq 4, 1 \leq t \leq 3\}$, and $C' = \{T_i : 1 \leq i \leq q\}$. From (ii) and the fact that $(\phi_j^z)^{-1}$ is onto E it follows that C' is a cover. From (i) and the fact that $(\phi_j^z)^{-1}$ is 1-1 it follows that C' is an exact cover. $\Box$

Lemma 3: Let O be a 3D object from the library defined in the proof of Theorem 3, and let O' be an object defined by 12 points from the picture in the proof of Theorem 3. If O can be mapped by translation rotation and stable perspective projection to O' then the mapping is with zero translation and rotation.

Proof: We use the following properties of perspective projection (see [7], Chapter 13): (a) Collinear 3D points are projected into collinear 2D points. (b) If the projection of parallel 3D lines is parallel 2D lines then the 3D lines are parallel to the image plane.

Let O be generated by $e_{i_1}, e_{i_2}, e_{i_3}$. Let L_j be the 3D line of the rotated and translated points $\phi_j^z(e_{i_1})$, $\phi_j^z(e_{i_2})$, $\phi_j^z(e_{i_3})$ for $1 \leq j \leq 4$, so that L_1 is parallel to L_2 and L_3 is parallel to L_4. Let u_j^1, u_j^2, u_j^3 be the points of O' that are mapped to $\phi_j^z(e_{i_1})$, $\phi_j^z(e_{i_2})$, $\phi_j^z(e_{i_3})$ respectively, then u_j^1, u_j^2, u_j^3 are collinear for $1 \leq j \leq 4$, and since the projection is stable, the 4 triplets are on 4 different lines in the picture. The picture has exactly four lines with at least 3 points. These lines are: $y = 0$, $y = m^3$, $x = -1$, and $x = m^3 + 1$. Therefore, the 4 triplets come from these 4 lines.

Let l_j be the projection of L_j for $1 \leq j \leq 4$. l_1 intersects with two lines from $\{l_2, l_3, l_4\}$, and is parallel to the third. Since L_1 intersects with L_3 and L_4, l_1 intersects with l_3, l_4, and is parallel to l_2. Thus, we have two parallel lines L_1, L_2 that are projected into parallel lines. Therefore, both L_1 and L_2 must be parallel to the image plane; let Z_1 and Z_2 be their depth. From the same arguments the lines L_3, L_4 are parallel to the image plane; let Z_3, Z_4 be their depth respectively. But since L_3 intersects with both L_1 and L_2 we have $Z_1 = Z_2 = Z_3 = Z_4$.

We conclude that all the points of the translated and rotated object O have the same distance from the image plane. From Equation (5) it follows that in this case the distance from the image plane has the effect of scaling the object. Thus, Lemma 3 follows from Lemma 2 when applied to the library of objects defined by $\phi_j(e_{i_1})$, $\phi_j(e_{i_2})$, $\phi_j(e_{i_3})$ and the 6 points u_j^1, u_j^2, u_j^3 for $1 \leq j \leq 2$. $\square$

6 Implications

In this section we identify constraints that can potentially simplify model based recognition, and other constraints that leave the problem NP-complete.

Local features other than a point: With no additional structure this can only make the problem more difficult. However, with additional structure of the local features the problem may become polynomial. For example, straight lines may have an additional constraint that their ends meet (see Figure 1).

Occlusion: Without additional structure this can only make the problem more difficult. However, with additional constraints such as convexity this makes our NP-completeness proofs inapplicable, so that it may potentially simplify the problem.

A small number of feature points: If each object is described by 2 points the problem is polynomially solvable by matching techniques.

A large number of feature points: Without additional structure this can only make the problem more difficult. However, if it is assumed that small subsets of these points determine a unique object from the library then the problem is polynomially solvable. (This is the essential assumption in geometric hashing [9]).

Almost distinct subsets: If the distance between every pair of feature points uniquely determines two (or less) objects, the problem is polynomially solvable. If this distance determines three (or more) objects the problem is still NP-complete. This follows from the comment in the definition of X3C.

Dimensionality: Notice that the results of Theorem 1 hold also for translation and rotation in 2 and 3 dimensions. Similarly, the results of Theorem 2 hold also for 3 dimensions.

7 Concluding remarks

We have shown that the problem of model based recognition is NP-complete. Thus, there is little hope for a performance guaranteed algorithm that can solve the problem efficiently. However, it is still possible that easy sub-classes of the problem can be characterized by additional structure of the modeled objects (e.g., convexity) and the way they are viewed (e.g., occlusion). Our results can help determine what constraints are potentially useful.

References

[1] P. J. Besl and R. C. Jain. Three-dimensional object recognition. *Computing Surveys*, 17(1):75–145, 1985.

[2] T. O. Binford. Survey of model-based image analysis systems. *International Journal of Robotics Research*, 1(1):18–64, 1982.

[3] R. T. Chin and C. R. Dyer. Model-based recognition in robot vision. *ACM Computing Surveys*, 18(1):67–108, 1986.

[4] M. R. Garey and D. S. Johnson. *Computers and Intractability. A Guide to the Theory Of NP-completeness.* Freeman, 1979.

[5] W. E. Grimson. The combinatorics of object recognition in cluttered environments using constrained search. In *Proceedings of the Second International Conference on Computer Vision (ICCV'88)*, pages 218–227, December 1988.

[6] W. E. L. Grimson. The combinatorics of local constraints in model-based recognition and localization from sparse data. *Journal of the ACM*, 33(4):658–686, 1986.

[7] R. M. Haralick and L. G. Shapiro. *Computer and Robot Vision.* Addison-Wesley, 1991.

[8] L. M. Kirousis and C. H. Papadimitriou. The complexity of recognizing polyhedral scenes. *Journal of Computer and System Sciences*, 37:14–38, 1988.

[9] Y. Lamdan and H. J. Wolfson. Geometric hashing: A general and efficient model-based recognition scheme. In *Proceedings of the Second International Conference on Computer Vision (ICCV'88)*, pages 238–249, December 1988.

[10] J. K. Tsotsos. Analyzing vision at the complexity level. *Behavioral and Brain Sciences*, 13:423–469, 1990.

[11] W. M. Wells III. A statistical approach to model matching. *Proc. SPIE Conf. on Intelligent Robots and Computer Vision IX: Algorithms and Techniques*, Boston, Nov. 1990.