Computational Limitations of Model-Based Recognition

Haim Schweitzer*
University of Texas at Dallas, P.O. Box 830688, Richardson, Texas 75083-0688

Sanjeev R. Kulkarni†
Department of Electrical Engineering, Princeton University, Princeton, New Jersey 08544

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 to avoid the inherent computational difficulties. © 1998 John Wiley & Sons, Inc.

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 many possible shapes of pens as such. 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 cataloged in a library. The problem of model-based recognition can be informally described in the following way: Given a library of modeled objects and 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 Refs. 3–5; more recent studies include Refs. 8 and 13. The standard computational approach is to represent the modeled objects and

*Author to whom correspondence should be addressed. E-mail: haim@utdallas.edu.
†E-mail: kulkarni@ee.princeton.edu.

INTERNATIONAL JOURNAL OF INTELLIGENT SYSTEMS, VOL. 13, 431–443 (1998)
 CCC 0884-8173/98/050431-13

the data in terms of discrete features so that the recognition can be solved as a search problem. The results of previous work show that by applying rigidity constraints in various ways, model-based recognition can be efficiently applied to recognize objects even from partial views and in the presence of nonmalicious 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 solvable subproblems. In the case of model-based recognition this corresponds to exploiting additional structure of the modeled objects and the way they are viewed. For more information on the theory of NP-completeness, see Ref. 6. For applications of NP-completeness results to vision tasks, see Refs. 14 and 12.

The results of this paper can be used to determine constraints that may simplify 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.

1.1. Comparison with Previous Work

Grimson[8,7] analyzed the complexity of model-based recognition for the special (important) case of tree search. His results show that if all of the data are known to have come from a single object, then the expected amount of search is a low order polynomial (in the number of model features). If, on the other hand, the object may not be present, the complexity becomes exponential (in the number of model features).

Grimson's results seem to imply that the search is always easy when objects can be modeled by a small number of features. If objects are modeled by many features, but they can be uniquely determined from a few, one can apply the "hypothesize and test" technique, which is the basic idea underlying alignment methods,[10] Hough transforms,[11,2] and geometric hashing.[13]

The other relevant complexity parameter (besides the number of features) is the library size. Previous work (e.g., Refs. 8, 7, 10, and 13) considered cases in which the library can be quite large and described implementations where the

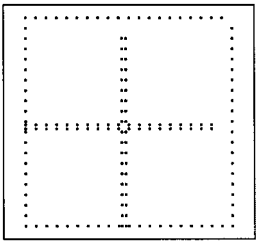

Figure 1. Different possible interpretations of sensed data (see text).

search is reduced to multiple search of single objects. The standard approach when the library is large is to sequentially test each model in the library for possible interpretations in the data. Once a correct interpretation is found, the data corresponding to it can be removed from consideration and the process is repeated on the remaining data. This "peel-off-one-at-a-time" technique gives complexity that is at most quadratic in the library size.

The main argument that we make in this paper is that the peel-off-one-at-a-time technique may not give the right interpretation of the data. In fact, the complexity of the search is most likely exponential in the library size even if the number of modeled features is a small constant. A situation where the straightforward peel-off approach may not work is illustrated in the following example. Let L be a library with a single modeled object, *a square*. The objects to be identified are squares and the sensed data are given as points along edges of the square. Consider the case shown in Figure 1. What appears to be the most dominant object is the outer square. If it is first detected and its points are removed, the rest of the points cannot be reliably interpreted as squares. A more reasonable interpretation that explains the entire data is the interpretation of four inner squares.

1.2. Paper Organization

The paper is organized as follows. Precise definitions for an object, a library, and a picture of sensed data are given in Section 2. Section 3 discusses the case in which the picture contains shifted and rotated objects from the library. It is shown that worst case exponential complexity is likely even if each object can be uniquely characterized by three points. Similar results for the more realistic cases in which the viewed objects can be shifted, rotated, and scaled objects from the library are presented in Section 3. The worst case exponential complexity proof requires that each object is modeled by six points. In most real-world situations the viewed objects are obtained as perspective projections of three-dimensional objects. Unfortunately, the general case of perspective projections includes extreme cases that may destroy or create geometric relations. The results of Section 4 generalize the results of Section 3 to perspective projections under the assumption that these projections do not accidentally create geometric relations. Some implications of the results are discussed in Section 6.

2. PRELIMINARY DEFINITIONS

We consider situations in which objects can be modeled in terms of sets of local features. A local feature is a simple geometric shape, and an object is described by a set of local features and their location in space. (The results of this paper hold for arbitrary interpretations of "simple" and "local.") Commonly used features are points, lines, angles, etc. An example is shown in Figure 2, where a triangle is described in terms of (a) straight lines, (b) corners, and (c) points along its edges.

DEFINITION. *An object description by local features is a set of t pairs*:

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

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

DEFINITION. *A library is a set of object descriptions.*

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

Model-Based Object Recognition Problem. For a family of coordinate transformations Ψ, a library L, and a picture of sensed data P given as the set of pairs

$$P = \{\langle f_1, X_1 \rangle, \dots, \langle f_m, X_m \rangle\}$$

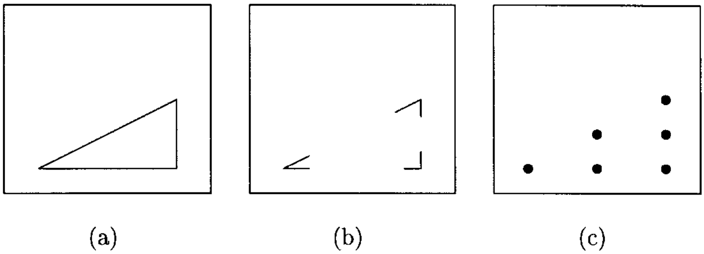

Figure 2. Examples of local features describing a triangle.

determine a disjoint partition of P into objects from L, i.e., subsets $O_1, \dots, O_q$ such that:

(i) For $i \neq j$ $O_i \cap O_j = \varnothing$.
(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 model-based recognition under translations, rotations, and (stable) perspective projections is NP-complete. The proofs are based on a reduction from *exact cover by three sets* (X3C) that is known to be NP-complete (see Ref. 6, p. 221).

Exact Cover by Three-Sets Problem. The following exact cover by three sets problem, referred to as X3C, is NP-complete:

Instance: A set E of m elements and a collection C of three-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. The related *exact cover by two-sets* problem is solvable in polynomial time.

3. THE CASE OF TRANSLATION AND ROTATION

In this section we analyze the complexity of recognizing objects that are assumed to be translated and rotated models from a given library.

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 library object is described by three 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, where C is a collection of three-element subsets of the m elements $e_1, \dots, e_m \in E$. We begin by constructing a picture P of m points $p_1, \dots, p_m$ 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_2, the point p_3 is at distance $m^2 + 2$ from p_2, etc. See the illustration in Figure 3. Let ϕ: $E \to P$ denote the mapping of elements in E to points in P. For $1 \leq i \leq m$ we have

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

Clearly, ϕ is $1-1$ and onto, so that the inverse mapping is well defined. We now create the library L from the three-element subsets in C. For a three-set

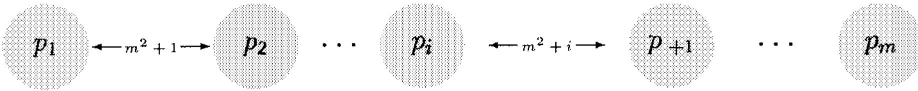

Figure 3. The picture in the proof of Theorem 1.

composed of the elements $e_\alpha, e_\beta, e_\gamma$ we add to L an object described by the three 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 4.

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 following proof makes use of Lemma 1 which is proved at the end of this section.

Let $C' \subset C$ be an exact cover of E. For each triplet $\{e_{i_1}, e_{i_2}, e_{i_3}\} \in C'$ define

$$O_i = \{\phi(e_{i_1}), \phi(e_{i_2}), \phi(e_{i_3})\}$$

so that $O_i \in L$. Since C' is a cover of E and ϕ is onto, we have $P = \bigcup_i O_i$. Since C' is exact and ϕ is 1–1, we have $O_i \cap O_j = \emptyset$ for $i \neq j$.

Conversely, let Ψ be the family of coordinate translations and rotations, and assume for each $O_i \in L$ the existence of $\psi_i \in \Psi$ such that

(i) For $i \neq j$, $\psi_i(O_i) \cap \psi_j(O_j) = \emptyset$.
(ii) $P = \bigcup_i \psi_i(O_i)$.

From Lemma 1 it follows that ψ_i must be the identity transformation, i.e., $\psi_i(O_i) = O_i$, so that $\psi_i(O_i) \in L$. For $O_i = \{p_{i_1}, p_{i_2}, p_{i_3}\}$ define

$$T_i = (\phi^{-1}(p_{i_1}), \phi^{-1}(p_{i_2}), \phi^{-1}(p_{i_3})), \qquad C' = \{T_i\}.$$

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. ■

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 three points from the picture in the proof of Theorem 1. If O can be mapped by translation and rotation to O', then $O = O'$.*

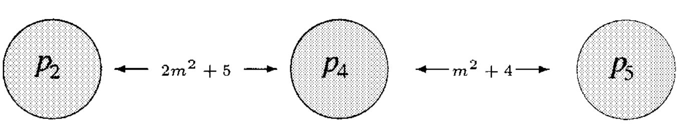

Figure 4. A typical object in the proof of Theorem 1.

Proof. Without loss of generality, let O be described by the points $p_{i_1}, p_{i_2}, p_{i_3}$ and let O' be described 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 one-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 Eq. (1) we have

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

Set $s(i, j) = (j(j-1) - i(i-1))/2$, 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 $1 \le i < j \le m$ and $|s(i_1, i_2) - s(j_1, j_2)| < m^2$, but since the right-hand side of Eq. (2) is divisible by m^2, it must equal 0 and we have

$$\begin{aligned} s(i_1, i_2) &= s(j_1, j_2) \\ j_2 - j_1 &= i_2 - i_1 \end{aligned} \qquad (3)$$

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 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°. However, 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

This section generalizes the results of Section 3 to transformations that may include scaling in addition to translation and rotation.

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 library object is described by six 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 two-dimensional 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 ϕ: $E \to P$ and θ: $E \to P'$:

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

See the illustration in Figure 5. We now create the library L from the three-element subsets of C. For $(e_\alpha, e_\beta, e_\gamma)$ we add to L an object described by the six 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 6.

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 following proof makes use of Lemma 2, which will be proved at the end of this section.

Let $C' \subset C$ be an exact cover of E. For each triplet $\{e_{i_1}, e_{i_2}, e_{i_3}\} \in C'$ 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 $O_i \in L$. Since C' is a cover of E, and ϕ, θ are onto P and P', respectively, $Q = P \cup P' = \bigcup_i O_i$. Since C' is exact and ϕ, θ are 1–1, $O_i \cap O_j = \varnothing$ for $i \neq j$.

Conversely, let Ψ be the family of coordinate translations, rotations, and scaling and assume for each $O_i \in L$ the existence $\psi_i \in \Psi$ such that:

(i) For $i \neq j$, $\psi_i(O_i) \cap \psi_j(O_j) = \varnothing$.
(ii) $Q = \bigcup_i \psi_i(O_i)$.

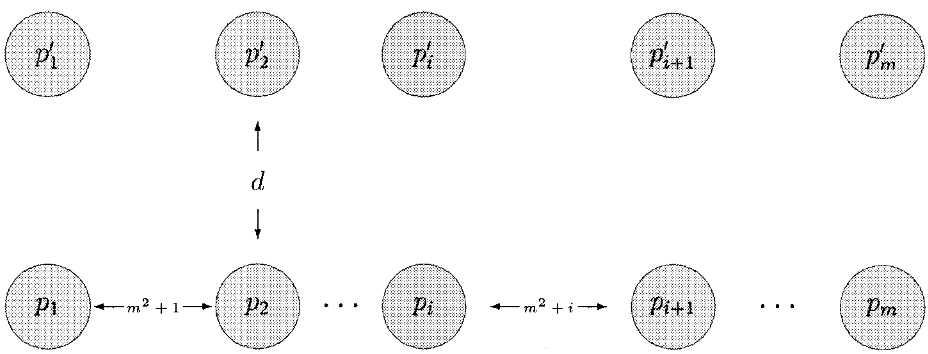

Figure 5. The picture in the proof of Theorem 2.

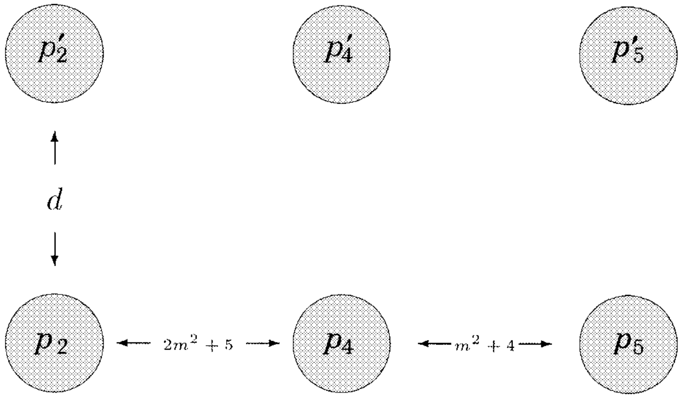

Figure 6. A typical object in the proof of Theorem 2.

From Lemma 2 it follows that ψ_i is the identity transformation, so that $\psi_i(O_i) \in L$. For $O_i = \{p_{i_1}, p_{i_2}, p_{i_3}, p'_{i_1}, p'_{i_2}, p'_{i_3}\}$, where it is assumed without loss of generality that $p_{i_1}, p_{i_2}, p_{i_3}$ have zero y coordinates, define

$$T_i = \left\{\phi^{-1}(p_{i_1}), \phi^{-1}(p_{i_2}), \phi^{-1}(p_{i_3})\right\}, \qquad C' = \{T_i\}.$$

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. ■

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 six 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 triplet lies on the line $y = d$. 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 $\{\theta(e_{i_1}), \theta(e_{i_2}), \theta(e_{i_3})\}$. ■

5. THE CASE OF PERSPECTIVE PROJECTION

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

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

This assumes the standard setup where the camera is at the origin, pointing directly up the Z axis. The reference frame is oriented as the image plane, which is located at distance f from the origin; see, e.g., Ref. 9. Unlike translation, rotation, and scaling, perspective projection may destroy geometric properties by merging lines and points, and create linear relations when they do not exist. 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 such cases it is necessary to impose some conditions on the projection. This was previously done, for example, in the work of Huffman and Waltz on the automatic interpretation of line drawings.[1] We define *stable* perspective projection views as follows:

DEFINITION. *A stable perspective projection of a scene satisfies the following properties*:

(i) *Distinct three-dimensional feature points are mapped into distinct two-dimensional feature points.*
(ii) *Noncollinear three-dimensional feature points are mapped into noncollinear two-dimensional feature points.*

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

THEOREM 3. *Let L be library of three-dimensional objects and let P be a two-dimensional picture given as a set of local features and their two-dimensional 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 library 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 two-dimensional picture $Q = P_1 \cup P_2 \cup P_3 \cup P_4$, where

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

$$\phi_1(e_i) = \text{a point at } x = (i-1)m^2 + i(i-1)/2, \qquad y = 0$$

$$\phi_2(e_i) = \text{a point at } x = (i-1)m^2 + i(i-1)/2, \qquad y = m^3$$

$$\phi_3(e_i) = \text{a point at } y = (i-1)m^2 + i(i-1)/2, \qquad x = -1$$

$$\phi_4(e_i) = \text{a point at } y = (i-1)m^2 + i(i-1)/2, \qquad x = m^3 + 1$$

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

We now create the library L from the three-element subsets of C. For $(e_{i_1}, e_{i_2}, e_{i_3})$ we add to L an object described by the 12 three-dimensional points $\{\phi_j^z(e_{i_t}); 1 \le j \le 4, 1 \le t \le 3\}$, where

$$\phi_1^z(e_i) = \text{a point at } X = (i-1)m^2 + i(i-1)/2, \qquad Y = 0, \qquad Z = f$$

$$\phi_2^z(e_i) = \text{a point at } X = (i-1)m^2 + i(i-1)/2, \qquad Y = m^3, \qquad Z = f$$

$$\phi_3^z(e_i) = \text{a point at } Y = (i-1)m^2 + i(i-1)/2, \qquad X = -1, \qquad Z = f$$

$$\phi_4^z(e_i) = \text{a point at } Y = (i-1)m^2 + i(i-1)/2, \qquad X = m^3 + 1, \qquad Z = f$$

Observe that the perspective projection $\pi(\phi_j^z(e_i)) = \phi_j(e_i)$ for $1 \le j \le 4$. It remains to show that the picture 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 following proof makes use of Lemma 3, which will be proved at the end of this section.

Let $C' \subset C$ be an exact cover of E. For each triplet $\{e_{i_1}, e_{i_2}, e_{i_3}\} \in C'$ define O_i as the three-dimensional object described by the 12 three-dimensional points $\{\phi_j^z(e_{i_t}); 1 \le j \le 4, 1 \le t \le 3\}$, so that $O_i \in L$. Since C' is a cover of E, and ϕ_j are onto P_j, respectively, $Q = \bigcup_i \pi(O_i)$. Since C' is exact and ϕ_j are 1–1, $\pi(O_i) \cap \pi(O_j) = \varnothing$ for $i \ne j$.

Conversely, let Ψ be the family of coordinate translations and rotations, and assume for each $O_i \in L$ the existence of $\psi_i \in \Psi$ such that

(i) For $i \ne j$, $\pi(\psi_i(O_i)) \cap \pi(\psi_j(O_j)) = \varnothing$.
(ii) $Q = \bigcup_i \pi(\psi_i(O_i))$.

From (ii) and Lemma 3 it follows that ψ_i is the identity transformation, so that $\psi_i(O_i) \in L$. Set $O_i = \{p_{i_1}^j, p_{i_2}^j, p_{i_3}^j\}$ for $1 \le j \le 4$, where we assume without loss of generality that $p_{i_t}^j$ were generated by ϕ_j^z. Define

$$T_i = \left\{ \left(\phi_j^z\right)^{-1}\left(p_{i_t}^j\right) : 1 \le j \le 4, 1 \le t \le 3 \right\}, \qquad C' = \{T_i\}.$$

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. ■

LEMMA 3. *Let O be a three-dimensional 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 the identity mapping.*

Proof. We use the following properties of perspective projection (see, e.g., Ref. 9, Chap. 13):

(a) Collinear three-dimensional points are projected into collinear two-dimensional points.
(b) If the projection of parallel three-dimensional lines is parallel two-dimensional lines, then the three-dimensional 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 three-dimensional 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 $1 \le j \le 4$, 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 $1 \le j \le 4$, and since the projection is stable, the four triplets are on four different lines in the picture. The picture has exactly four lines with at least three points. These lines are $y = 0$, $y = m^3$, $x = -1$, and $x = m^3 + 1$. Therefore, the four triplets come from these four lines.

Let l_j be the projection of L_j for $1 \le j \le 4$, 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. 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 points of the translated and rotated object O have the same distance from the image plane. From Eq. (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 six points u_j^1, u_j^2, u_j^3 for $1 \le j \le 2$. ■

6. IMPLICATIONS

In this section we briefly mention 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 be solvable in polynomial time. For example, straight lines may have an additional constraint that their ends meet (see Fig. 2).

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 library object can be uniquely described by two points, the problem of model-based recognition can be solved in polynomial time 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 solvable in polynomial time. (This is the essential assumption in geometric hashing.[13])

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 also hold for translation and rotation in two and three dimensions. Similarly, the results of Theorem 2 also hold for three dimensions.

7. CONCLUDING REMARKS

We have shown that the problem of model-based recognition is NP-complete. Thus, there is little hope for performance guaranteed algorithms that can solve the problem efficiently. However, it may still be possible that easy subclasses 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.

The work of the first author was supported in part by the National Science Foundation under Grant IRI-9309135.

References

1. D.H. Ballard and C.M. Brown, *Computer Vision*, Prentice-Hall, Englewood Cliffs, NJ, 1982.
2. J.R. Bergen and H. Shvaytser, "A probabilistic algorithm for computing Hough transforms," *J. Algorithms*, **12**(4), 639–656 (1991).
3. P.J. Besl and R.C. Jain, "Three-dimensional object recognition," *Comput. Surveys*, **17**(1), 75–145 (1985).
4. T.O. Binford, "Survey of model-based image analysis systems," *Int. J. Robotics Res.*, **1**(1), 18–64 (1982).
5. R.T. Chin and C.R. Dyer, "Model-based recognition in robot vision," *ACM Comput. Surveys*, **18**(1), 67–108 (1986).
6. M.R. Garey and D.S. Johnson, *Computers and Intractability. A Guide to the Theory of NP-Completeness*, Freeman, San Francisco, 1979.
7. W.E.L. Grimson, "The effect of indexing on the complexity of object recognition," in *Proceedings Third Int. Conf. on Computer Vision* (*ICCV'90*), pp. 644–651, 1990.
8. W.E.L. Grimson, *Object Recognition by Computer: The Role of Geometric Constraints*, The MIT Press, Cambridge, MA, 1990.
9. R.M. Haralick and L.G. Shapiro, *Computer and Robot Vision*, Addison-Wesley, Reading, MA, 1992.
10. D.P. Huttenlocher and S. Ullman, "Recognizing solid objects by alignment with an image," *Int. J. Computer Vision*, **5**(2), 195–212 (1990).
11. J. Illingworth and J. Kittler, "A survey of the Hough transform," *Computer Vision, Graphics, Image Process.*, 44, 87–116 (1988).
12. L.M. Kirousis and C.H. Papadimitriou, "The complexity of recognizing polyhedral scenes," *J. Computer Syst. Sci.* **37**, 14–38 (1988).
13. Y. Lamdan and H.J. Wolfson, "Geometric hashing: A general and efficient model-based recognition scheme," in *Proc. Second Int. Conf. on Computer Vision* (*ICCV'88*), pp. 238–249, 1988.
14. J.K. Tsotsos, "Analyzing vision at the complexity level," *Behavioral Brain Sci.*, **13,** 423–469 (1990).