

Recognition Using Region Correspondences

Ronen Basri

Department of Applied Math.
The Weizmann Inst. of Science
Rehovot, 76100, Israel

David Jacobs

NEC Research Institute
4 Independence Way
Princeton, NJ 08540, USA

Abstract

A central problem in object recognition is to determine the transformation that relates the model to the image, given some partial correspondence between the two. This is useful in determining whether an object is present in an image, and if so, determining where the object is. We present a novel method of solving this problem that uses region information. In our approach the model is divided into volumes, and the image is divided into regions. Given a match between subsets of volumes and regions (without any explicit correspondence between different pieces of the regions) the alignment transformation is computed. The method applies to planar objects under similarity, affine, and projective transformations and to projections of 3-D objects undergoing affine and projective transformations.

1 Introduction

A fundamental problem in recognition is *pose estimation*. Given a correspondence between some portions of an object model and some portions of an image, determine the transformation that relates the model with the image. This is obviously essential if we wish to determine the position of objects in the world from their appearance in an image. Also, to recognize objects we frequently seek to eliminate the effects of viewpoint by bringing the model and the image into alignment.

We present a novel method for determining the pose of a known object based on matching portions of a known model, and some (possibly occluded) areas of the image. Our method finds a model pose that will project these portions of the model onto the corresponding image areas, without requiring knowledge of the correspondence between specific points in the model and image. An example is shown in Figure 1. We show that in general a small number of region correspondences determine the correct pose of the object uniquely and present a number of experiments.

The novelty of our method lies in its use of region information to determine pose. Current approaches to pose determination include global methods (e.g., [12, 21]) that distinguish shapes and determine pose using properties of an entire shape such as its moments, and local methods, which match local model and image features and use these matches to determine pose (e.g., [10, 15, 22, 1, 19, 3]). Global methods have the advantage of efficiency, since processing the image can be carried out independently of the model.

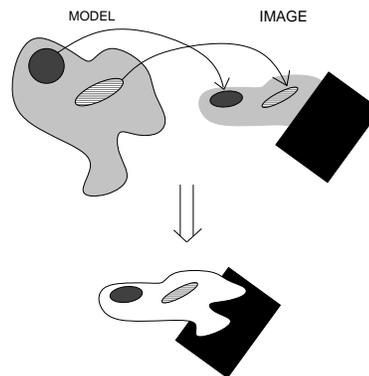


Figure 1: An example of region-based pose determination. The two matched regions determine the pose under an affine transformation.

However, such methods require a good segmentation of the object, making them sensitive to occlusion, and in particular to self occlusion. This makes these methods unsuitable for recognizing 3D objects from single 2D images, or to recognition in cluttered scenes.

Local methods are resistant to partial occlusion or to segmentation failures, since they can make use of fragmented data. They have encountered some significant limitations, however. First, it has proven extremely difficult to reliably locate local features in images of 3-D objects. Second, local methods can be computationally expensive, requiring extensive search to find correspondences.

Techniques have been applied to the recognition of curved objects that use algebraic descriptions of extended curves (e.g., [18, 23, 11, 17]). This approach can potentially overcome some of the difficulties of isolating local features reliably. However, the need for extended contour segments can make this approach vulnerable to partial occlusion, and it can be difficult to robustly extract an algebraic description of a curved contour fragment.

We believe that a fundamental difficulty with these approaches is that the position of local features does not adequately capture the appearance of an object. As an example, in Figure 2 we show two different polygonal shapes. The locations and number of the vertices of these polygons differ considerably, while the overall shapes of the objects are quite similar. In gen-

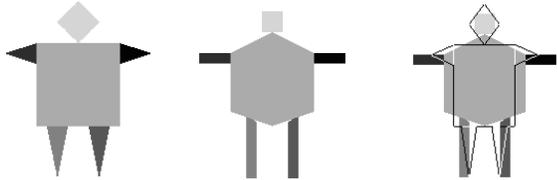


Figure 2: Two polygons that differ in features yet share the same overall structure. On the right, an alignment of the two figures using our method.

eral, small changes in the surface of an object can have a large effect on the location of features derived from the object contour, or on the algebraic description of the contour. This explains why local features are an inherently unstable way of describing many objects. It also explains why local features are poorly suited for comparing different objects that are instances of the same class of objects. Two different chairs, for example, may be quite similar in having arms, legs, a seat and a back of roughly similar shape and position, but still produce corners in totally different positions.

For this reason we propose a recognition method based on matching volumes of an object to regions of an image. In our approach the model is divided into convex volumes, and the image is divided into convex regions. (Convexity is not a limitation, however, since concave entities can be replaced by their convex hulls.) Given a match between subsets of regions and volumes the alignment transformation is computed. Figure 2 shows that our approach achieves a rough alignment of the two polygons by matching the regions themselves, rather than local boundary features. In this paper we will focus primarily on developing this method for the recognition of a planar model in a 2-D image, taken from an arbitrary 3-D viewpoint. In [5] we show that the method can be extended to the recognition of a 3-D object in a 2-D image. In summary, our method has the following advantages: region boundaries need not be precisely localized; the combinatorics of matching can be reduced by using color and texture information identified with regions; the method handles smooth and curved objects, regions and local features, all in the same framework; the method allows for partial occlusion.

2 Problem Definition

Below we consider the following problem. Given a set of model volumes $V_1, \dots, V_k \subset \mathcal{R}^d$, $d \in \{2, 3\}$, and a corresponding set of image regions $R_1, \dots, R_k \subset \mathcal{R}^2$ determine the transformation $T \in \mathcal{T}$ that maps every volume V_i to its corresponding region R_i ($1 \leq i \leq k$). In the 2D case the allowed set of transformations, \mathcal{T} , is either the similarity (rotation, translation, and uniform scaling), affine (linear and translation), or projective transformations. In the 3D case we consider affine transformations followed by either an orthographic or perspective projection.

Throughout the paper we assume that the volumes and regions are all closed and convex. The solutions we propose handle non-convex regions by replacing them with their convex hulls. Points and line segments fall naturally into this formulation as they form convex sets.

The problem of determining the transformation that maps a set of model volumes to their corresponding image regions is generally non-convex. That is, the set of feasible transformations need not be convex, or even connected. Consider for example the case of a model square matched to an image containing an identical square. Matching the model exactly to the image can be performed in four ways corresponding to rotating the model square so as to bring any of its four sides to the top. Obviously, no intermediate transformation provides a solution to this matching problem.

To handle this problem, we distinguish between two sets of constraints.

Forward constraints: every model point $\vec{p} \in V_i$ should project inside the region R_i (that is, $TV_i \subseteq R_i$).

Backward constraints: every image point $\vec{q} \in R_i$ is the projection of some volume point $\vec{p} \in V_i$ (that is, $TV_i \supseteq R_i$).

Below we consider the problem of computing the transformation T , of a family of transformations \mathcal{T} , that is consistent with either the forward constraints or the backward constraints. We see that individually each set of constraints produces a convex set of feasible transformations, although the combination of the two does not. This leads to efficient solutions when we use only one set of constraints. Moreover, the backward constraints express precisely our state of knowledge should we wish to allow for arbitrary amounts of occlusion of image regions.

3 The 2D problem

In this section we show that for similarity, affine, and projective transformations, the one-way constraints can be formulated as a set of linear inequalities with the transformation parameters as the unknowns. Determining the transformation that relates the model to the image is equivalent to finding a linear discriminant function. In particular, the solution can be obtained by applying a linear program. We show that a unique solution to the one-way matching problem generally exists for as few as two distinct regions.

3.1 One-way constraints

We denote a point in model space by $\vec{p} = (x, y)$ and in image space by $\vec{q} = (u, v)$. When $\vec{q} = T(\vec{p})$ we denote $u = T_u(\vec{p})$ and $v = T_v(\vec{p})$. We begin by defining the one-way constraints.

We formulate the forward constraints as follows. Given a convex model volume V and a corresponding convex image region R we want to find a transformation T that maps V inside R . Note that both V and R might in particular be simply points or line segments. Since R is convex, there exists a set of lines L_R bounding R from all directions such that for every point $\vec{q} \in R$ and for every line $l \in L_R$ we can write

$$l(\vec{q}) \geq 0. \quad (1)$$

The constraint $TV \subseteq R$ can be written as follows. Every point $\vec{p} \in V$ should be mapped by T to some point $\vec{q} \in R$, and so

$$l(T\vec{p}) \geq 0. \quad (2)$$

The set of forward constraints consists of all such constraints obtained for all pairs of bounding lines $l \in L_R$ and model points $\vec{p} \in V$. (Therefore, the set of forward constraints is homomorphic to $L_R \times V$.) When several model volumes V_1, \dots, V_k and corresponding image regions R_1, \dots, R_k are given the set of forward constraints is the union of sets of constraints for each pair of corresponding model V_i and region R_i . The back constraints are obtained in the same way by interchanging model volumes with image regions. Below we derive the constraints (2) explicitly for both the forward and backward cases allowing either similarity, affine, or projective transformations.

Forward constraints. Let $\vec{p} = (x, y) \in V$ be a model point, and let $Au + Bv + C \geq 0$ be a half space containing R . The forward constraints take the form

$$AT_u(\vec{p}) + BT_v(\vec{p}) + C \geq 0. \quad (3)$$

where the unknowns are the parameters of the transformation T .

Similarity transformation. Suppose T is a similarity transformation, we can write T in the following form

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} c \\ d \end{pmatrix}. \quad (4)$$

Denote $\vec{w}^T = (a, b, c, d)$ and $\vec{g}^T = (Ax + By, Ay - Bx, A, B)$. Then, from Equations (3) and (4) we obtain

$$\vec{g}^T \vec{w} \geq -C. \quad (5)$$

Baird pointed out that linear bounds on the location of a transformed model point leads to linear constraints on the feasible transformations [4] (see also [8, 6]).

Affine transformation. If T is an affine transformation, T is given in the form

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix}. \quad (6)$$

We can again rewrite the forward constraint as in Eq. (5), letting $\vec{w}^T = (a, b, c, d, e, f)$ and $\vec{g}^T = (Ax, Ay, Bx, By, A, B)$.

Projective transformation. While convexity is not preserved under arbitrary projective transformations of planar shapes, it is preserved under perspective projections. Therefore, in real images, convex model volumes will always appear as convex image regions. In the projective case, T , can be expressed in the form

$$\alpha \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} a & b & e \\ c & d & f \\ g & h & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (7)$$

for some arbitrary scalar factor α . Thus,

$$u = \frac{ax + by + e}{gx + hy + 1} \quad v = \frac{cx + dy + f}{gx + hy + 1}. \quad (8)$$

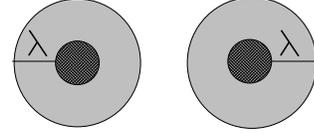


Figure 3: The dark circles are positioned by the similarity transformation that maximizes λ relative to the larger, shaded circles.

As we require the image plane to separate the object from the center of projection we can assume WLOG that the depth coordinate, $gx + hy + 1$, is positive for all points. Imposing the constraint (3) and denoting $\vec{w}^T = (a, b, c, d, e, f, g, h)$ and $\vec{g}^T = (Ax, Ay, Bx, By, A, B, Cx, Cy)$ we again obtain the constraint as in Eq. (5).

Backward constraints. In the 2D case models and images are interchangeable, and so the backward constraints can be defined in the same way as the forward constraints. Again, for each type of projection, the constraints can be written as in Eq. (5). But this time the vector of unknowns, \vec{w} , represents the image-to-model transformation, which is the inverse of the transformation solved for by the forward constraints.

Solving for the transformation using the backward constraints alone is particularly useful in the case of occlusion. Image regions that are partly occluded lie inside the corresponding model volumes (after the model and the image are brought into alignment), but the inclusion may be strict due to the occlusion.

3.2 Solving the one-way constraints

The one-way problem under affine, similarity, or projective transformations introduces a set of linear constraints in the transformation parameters. In the forward problem the set of constraints contains one constraint for every point in the model volumes and for every tangent line to the image regions. In the backward problem the model and image change roles. The number of constraints for a curved object is therefore infinite. For polygonal volumes and regions the number of independent constraints is finite. The system of constraints in this case is completely defined by the vertices of the model volumes and the sides of the image regions, and the rest of the constraints are redundant. In the curved case we will want to sample the set of constraints. The issue of sampling is addressed in [5].

We can solve a finite set of linear constraints

$$\vec{g}_i^T \vec{w} \geq c_i, \quad i = 1, \dots, n \quad (9)$$

using linear programming. To generate a linear program a linear objective function should be specified. A common way of defining such a linear program is by introducing an additional unknown, λ , and solving

$$\max \lambda \quad s. t. \quad \vec{g}_i^T \vec{w} \geq c_i + \lambda, \quad i = 1, \dots, n \quad (10)$$

A solution to (9) exists if and only if a solution to (10) with $\lambda \geq 0$ exists.

When $\lambda \geq 0$ its value represents the minimal distance of a point to any line bounding the region (Figure 3). Maximizing λ amounts to attempting to contract the model volume inside the image region as much as possible. When $\lambda < 0$ this attempt fails. In this case any model point that violates the constraints is mapped to a distance of no more than $|\lambda|$ from its target regions. ($|\lambda|$ in this case represents a maximum norm, and so it is related to the Hausdorff metric. For work on Hausdorff matching, see [13, 14, 2]).

Solving the system (10) may result in over-contraction. Consider, for example, the case of matching a single volume V to a single region R . Many transformations will shrink V to a small size and place it entirely inside R . The solution picked by Eq. 10 scales V down to a point, which is then translated to the point inside R furthest from any of its bounding tangent lines. Clearly, the case of matching one volume to one region cannot be solved by the one-way constraints alone. In what follows we prove that generally if the model contains two or more non-overlapping regions the solution is unique.

3.3 Uniqueness theorems

In this section we establish the conditions under which a one-way region matching problem has a unique solution. We assume that regions are produced by applying a transformation to a set of volumes. We consider under what circumstances a corresponding transformation that will project the volumes entirely inside the corresponding regions is uniquely determined. Note that in the absence of occlusion, whenever this transformation is unique, the inverse transformation found by the backward constraints will also be unique, since the forward and backward matching problems are identical.

We begin by proving Lemma 1 which establishes that the uniqueness of a one-way matching problem is determined by the model alone. If a model is non-degenerate a unique solution will be obtained when the model is matched to any of its images, while if the model is degenerate, multiple solutions will exist when the model is matched to any image of the object. The lemma states the following claim. The solution to a one-way matching problem under a certain group of transformations (similarity, affine, or projective) is unique if and only if there exists no transformation of that group (other than the trivial one) which projects the model volumes entirely inside themselves.

Using Lemma 1 we show that in the similarity case two distinct (non-intersecting) model volumes and their corresponding image regions determine the transformation uniquely. In the affine case we show that three volumes positioned such that no straight line passes through all three volumes determine the transformation uniquely. Then, we derive necessary and sufficient conditions for two volumes to determine a unique solution. Degenerate cases are analyzed in [5]. Similarly, in the projective case we show that three volumes positioned such that no straight line passes through all three volumes determine the transformation uniquely. We now turn to showing that uniqueness is dependent on the model alone.

Lemma 1: *Let $V_1, V_2, \dots, V_k \subseteq \mathcal{R}^2$ be k distinct (non-intersecting) volumes. Let \mathcal{T} be the group of similarity, affine, or projective transformations. Let $R_i = T(V_i) \subseteq \mathcal{R}^2$, $1 \leq i \leq k$ be k regions obtained from V_1, \dots, V_k by applying an invertible transformation $T \in \mathcal{T}$. Then, there exists a transformation $T' \neq T, T' \in \mathcal{T}$ such that $T'(V_i) \subseteq R_i$, $1 \leq i \leq k$, if and only if there exists a transformation $\tilde{T} \neq I, \tilde{T} \in \mathcal{T}$ (I denotes the identity transformation) such that $\tilde{T}(V_i) \subseteq V_i$ for all $1 \leq i \leq k$.*

Proof: Suppose there exists a transformation $\tilde{T} \neq I$ such that $\tilde{T}(V_i) \subseteq V_i$ for all $1 \leq i \leq k$. Let $T' = T\tilde{T}$. Clearly, $T' \neq T$ and $T'(V_i) \subseteq R_i$. Conversely, assume there exists a transformation $T' \neq T$ such that $T'(V_i) \subseteq R_i$. Let $\tilde{T} = T^{-1}T'$. Then $\tilde{T} \neq I$ and $\tilde{T}(V_i) \subseteq V_i$. Furthermore, since $\tilde{T} = T^{-1}T'$ then \tilde{T} belongs to same group as T and T' . \square

3.3.1 Similarity transformations

In this section we show that a similarity transformation is determined uniquely by two distinct volumes.

Theorem 2: *Let $V_1, V_2 \subseteq \mathcal{R}^2$ be two distinct convex closed volumes ($V_1 \cap V_2 = \emptyset$). Then, the solution to the one-way matching problem with these volumes as a model under a similarity transformation is unique.*

Proof: According to Lemma 1 the solution to the one-way matching problem is unique if and only if there exists no similarity transformation other than the trivial one that maps V_1 and V_2 to inside themselves. Let T be a similarity transformation such that $T(V_1) \subseteq V_1$ and $T(V_2) \subseteq V_2$. Since V_1 and V_2 are both closed and convex, and since T is a continuous transformation mapping these two volumes to inside themselves then, by Brouwer's fixed point theorem [9], there exist two points $\vec{p}_1 \in V_1$ and $\vec{p}_2 \in V_2$ that are fixed with respect to T , that is, $T(\vec{p}_i) = \vec{p}_i$ ($i=1,2$). (Note that $\vec{p}_1 \neq \vec{p}_2$ since V_1 and V_2 are distinct.) Two points determine a similarity transformation uniquely. Therefore, the identity transformation is the only similarity transformation that maps the two volumes to within themselves, and so $T = I$. \square

3.3.2 Affine transformations

Next we show that an affine transformation is uniquely determined by three volumes that cannot be traversed by any straight line. Later we derive a necessary and sufficient condition for two volumes to determine a unique solution.

Theorem 3: *Let $V_1, V_2, V_3 \subseteq \mathcal{R}^2$ be three distinct closed volumes such that there exists no straight line passing through all three volumes. Then, the solution to the one-way matching problem with these volumes as a model under an affine transformation is unique.*

Proof: Similar to Theorem 2, assume T is an affine transformation that maps the volumes to inside themselves. Then there exist three points that are

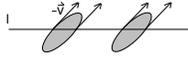


Figure 4: Two model volumes lead to non-unique affine transformations when a line l , exists such that the tangents at all intersection points are parallel. In this case, the volumes can contract towards l in the direction \vec{v} .

fixed with respect to T . Since no straight line passes through all three volumes the three fixed points are non-collinear, and so they determine the identity as the only affine transformation that maps the volumes to inside themselves. Therefore $T = I$. \square

We now turn to showing that the number of volumes required to determine the affine transformation uniquely is in general two.

Theorem 4: *Let $V_1, V_2 \subseteq \mathcal{R}^2$ be two distinct closed volumes. Then, the solution to the one-way matching problem with these volumes as a model under an affine transformation is not unique if and only if there exists a line l through V_1 and V_2 and a direction \vec{v} such that contracting V_1, V_2 in the direction \vec{v} toward l (denoted by $T_{l, \vec{v}}$) implies*

$$T_{l, \vec{v}}(V_i) \subset V_i \quad i = 1, 2.$$

(see Figure 4).

Proof: One direction is straightforward. Assume $T_{l, \vec{v}}$ contracts the volumes within themselves. $T = T_{l, \vec{v}}$ is itself an affine transformation (different from the identity transformation). To see this, let l be the x -axis, without loss of generality, and let $\vec{v} = (v_x, v_y)$. Then this affine transformation is given by:

$$T_{l, \vec{v}}(p) = \begin{pmatrix} 1 & v_x \\ 0 & 1 + v_y \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Conversely, assume the solution to the one-way matching problem is not unique. According to Lemma 1 there exists an affine transformation $T \neq I$ such that $T(V_i) \subset V_i$ ($i = 1, 2$). We next show that T is $T_{l, \vec{v}}$. Since \bar{T} maps the two volumes to within themselves there exist two points $\vec{p}_1 \in V_1$ and $\vec{p}_2 \in V_2$ that are fixed with respect to T , $T(\vec{p}_i) = \vec{p}_i$ ($i = 1, 2$). Since $V_1 \cap V_2 = \emptyset$, $\vec{p}_1 \neq \vec{p}_2$ and the points determine a line. This line is pointwise-fixed with respect to T ,

$$T(\vec{p}_1 + \alpha(\vec{p}_2 - \vec{p}_1)) = \vec{p}_1 + \alpha(\vec{p}_2 - \vec{p}_1)$$

for any scalar α . Denoting the fixed line by l , we now show that T represents a contraction in some direction \vec{v} toward l . Assume without the loss of generality that $\vec{p}_1 = \vec{0}$ and that l coincides with the X -axis, then T must have the form:

$$T(p) = \begin{pmatrix} 1 & a \\ 0 & b \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Denote the angle between \vec{v} and l by ψ , then contraction in a direction \vec{v} toward l is expressed by

$$(x, y) \longrightarrow (x + (s - 1)y \cot \psi, sy)$$

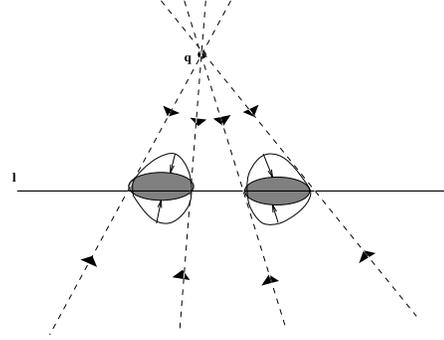


Figure 5: Two model volumes lead to non-unique projective transformations when a line l , exists such that the tangents at all intersection points meet at a single point q . In this case, the volumes can contract towards l in the directions emanating from q .

for some scalar $s < 1$. T represents such a contraction since we can set $s = b$ and $\psi = \cot^{-1} \frac{a}{b-1}$. \square

Theorem 4 above shows that any two non-intersecting regions provide a unique affine solution unless one can draw a line through the regions and contract the regions toward that line so that the regions would lie entirely inside themselves. For shapes in general position, such a line will not exist. An analysis of the degenerate cases is given in [5].

3.3.3 Projective transformations

Similar results extend to the projective case. Using the same techniques as in the similarity and the affine cases it is straightforward to show that four volumes such that no straight line passes through any three of the volumes determine the projective transformation uniquely. (Simply since four points generally determine a projective transformation.) Four, however, is not the minimal number of volumes that determine a unique solution. We are able to show:

Theorem 5: *Let $V_1, V_2, V_3 \subseteq \mathcal{R}^2$ be three closed volumes with non-zero areas such that there exists no straight line passing through all three volumes. Then, the solution to the one-way matching problem with these volumes as a model under a projective transformation is unique.*

Theorem 6: *Let $V_1, V_2 \subseteq \mathcal{R}^2$ be two distinct closed volumes with non-zero areas. Then, the solution to the one-way matching problem with these volumes as a model under a projective transformation is nonunique if and only if there exists a line l through V_1 and V_2 and a point q outside V_1, V_2 and l such that the following condition is met. Let p_i be any point at the intersection of V_i and l . Then the tangent line to V_i at the point p_i includes q . More informally, this will imply that contracting V_1 and V_2 in directions emanating from q toward l (denoted by $T_{l, q}$) implies*

$$T_{l, q}(V_i) \subset V_i \quad i = 1, 2.$$

(see Figure 5).

This theorem is the natural generalization of the two region case under affine transformations. In that case, a degeneracy occurs when the tangent lines are parallel (i.e., intersect at a point at infinity). In the projective case, a degeneracy occurs when the tangent lines intersect at any point in the plane.

The proof of these theorems is somewhat complex, and is given in [5].

3.3.4 Points and line segments

When applying our method we may wish to use points or line segments in addition to regions. By applying the results introduced in this section we can analyze what combinations of points and lines determine the transformation uniquely under a one-way matching problem. These combinations are specified in [5].

4 The 3D problem

It is readily shown that a system of forward constraints in the 3-D case can be solved in the same way such constraints are solved in the 2-D case, provided that we continue to use projection models without non-linear constraints. In [5] we provide details of this method for the case of 3-D to 2-D affine and projective transformations. It also contains proofs of the following uniqueness theorems:

Theorem 7: *There exist views of any pair of model volumes that lead to non-unique transformations, using the forward constraints.*

Theorem 8: *Applying the forward constraints to a set of volumes and regions leads to a non-unique solution only if:*

1. *There exists a plane P , which intersects each volume in two points.*
2. *For all points in the intersection of P and a volume, the direction of tangency to the projections of the volume are parallel. Note that if R_i does not have a smooth boundary, the tangent directions at the point of intersection may be undefined, but bounded. In this case, the possible ranges of the tangent vectors must intersect.*

This implies that four planar volumes that are not intersected by a plane always lead to a unique solution.

5 Experiments

To test the scheme we took pictures of a number of roughly planar objects. We first processed these images using Canny's edge detector [7]. We then constructed polygonal approximations to the edges using Pavlidis and Horowitz's [20] split-and-merge algorithm. The resulting polygons approximate the original edges to within two pixels. Then, we extracted the roughly convex structures using Jacobs's grouping system [16]. The matching between the regions was specified manually. Finally, the transformations relating these images were recovered using either the forward or backward solutions.

Figure 6 shows an image of a diskette used as a model, and Figure 7 shows the result of matching this model to another image of the diskette by solving for a similarity and for an affine transformation using all five regions. In this case the amount of affine distortion in the image is small, and so a good match was obtained in both cases. Figure 8 shows the result of matching the model to the same image using only two regions and the result of matching when two degenerate (with respect to an affine transformation) regions are used. These regions are degenerate because there exist four collinear points on their boundaries such that their tangent vectors are parallel. Notice the good match obtained in the similarity solution and the contraction produced in the affine solution.

Figure 9 demonstrates the performance of the system in the presence of partial occlusion. Notice that a good match was obtained when the backward constraints are used, whereas a contraction was obtained when the forward constraints are used. In this image, three of the five regions are occluded. Since the remaining two regions are degenerate by themselves, the partial information obtained from the occluded regions is essential to producing an accurate result.

Figure 10 shows the application of the projective method to an image of the diskette containing large perspective distortions. The match for this picture is significantly better than that obtained under the affine solution.

Figure 11 shows the application of the method to images of a magnet. It can be seen that a good match was obtained for these images, although some of the regions in the picture are not well localized.

Finally, Figure 12 shows two images of a book. Three regions were extracted from these images and used to determine the 2-D affine transformation that relates the two images.

The experiments demonstrate that our method obtains good results when applied to realistic objects. The system overcomes reasonable noise, in particular due to sparse sampling, and recovers the transformation successfully even in the presence of partial occlusion.

6 Conclusion

In spite of the success of model-based recognition techniques in many application areas, they still have significant weaknesses. Some of these weaknesses are due to the problem of representation. Most model-based techniques rely on a representation of objects in terms of local, precisely localizable features, or on algebraic descriptions of more extended portions of contours. While often quite valuable, these representations have the disadvantage that they describe the boundary of an object, not its internal shape. If one perturbs the boundary of an object a bit, one can completely alter the local features or algebraic curves that describe it, without changing the internal structure much. Our approach suggests a different way of representing objects for recognition. We represent and make use of the internal shape of objects, not just their boundary. And we suggest a way of making use of hybrid representations of objects that capture internal

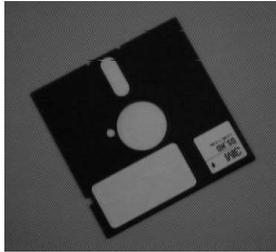


Figure 6: An image of a computer diskette used as a model.

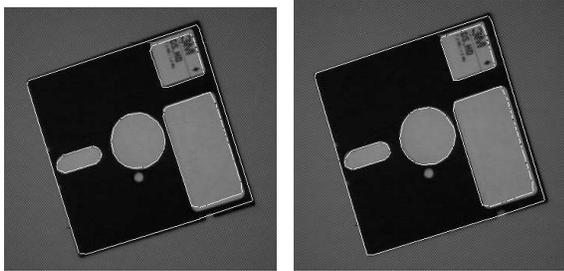


Figure 7: Matching the diskette model to a novel image of the diskette under similarity (left, $\lambda = -1.89$) and affine (right, $\lambda = -1.27$) transformations. In each case the model regions' positions are indicated by white lines, overlaid on top of the image.

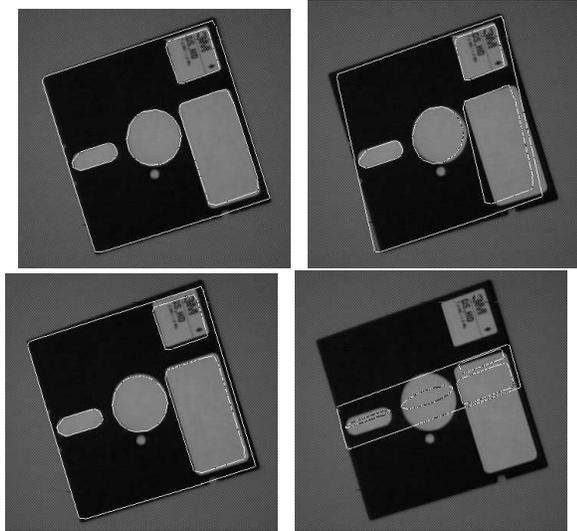


Figure 8: Top pictures: matching the diskette model to a novel image of the diskette using two regions only (the left and the upper right, left figure: similarity, $\lambda = -1.56$, right: affine, $\lambda = -0.69$). Bottom: the diskette model matched to a novel image of the diskette using two degenerate regions only (the left and the lower right, left figure: similarity, $\lambda = -1.15$, right: affine, $\lambda = -0.55$).

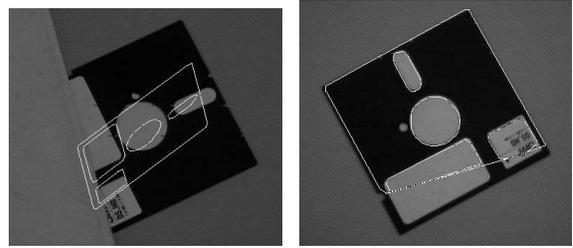


Figure 9: Matching the diskette model to a novel image that contains occlusion under an affine transformation using the forward (left figure, $\lambda = -12.59$) and the backward (right, $\lambda = -1.51$) constraints.

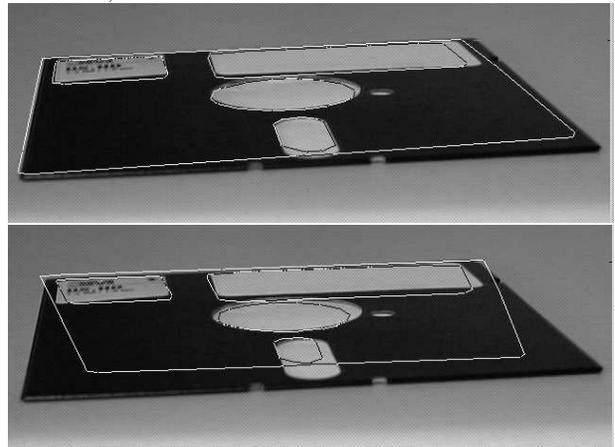


Figure 10: Matching the diskette model to an image containing relatively large perspective distortions under projective (left, $\lambda = -2.25$) and affine (right, $\lambda = -5.38$) transformations.

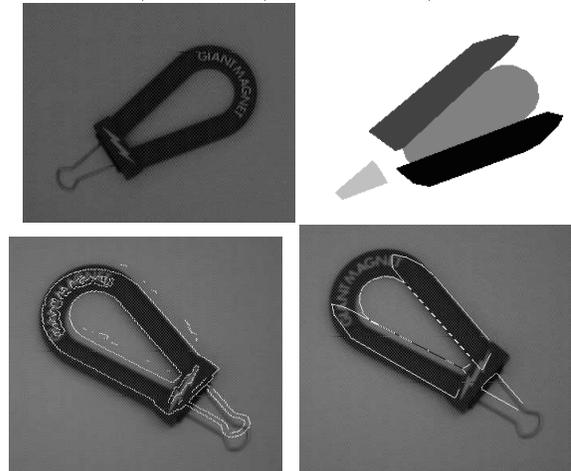


Figure 11: Matching a model of a magnet to a novel image: the model (top left), the regions extracted from the model (top right), the match (under affine transformation, bottom left, $\lambda = -3.46$), and the overlaid regions (bottom right).

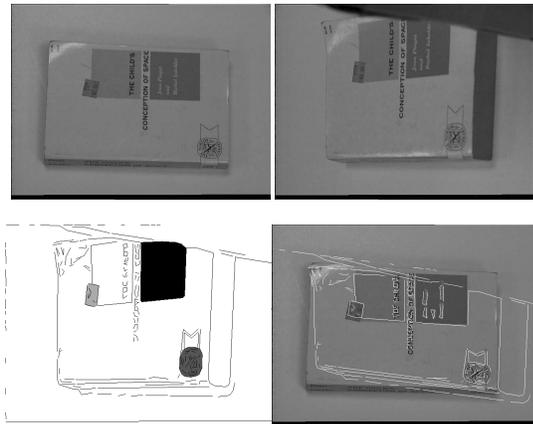


Figure 12: Top: a model image (left) and an occluded image of a book (right). Bottom: the three regions extracted (left, the regions are shaded) and the match obtained (right, $\lambda = -3.82$).

shape and local boundary structure when available.

Perhaps what is most novel about this approach is the weaker requirements that it makes on correspondence, compared to previous approaches. If all that we really know is that some portion of the image, of whatever extent, was produced by some specific portion of the model, our method allows us to make use of this information as well. Our method should therefore be seen as an extension to past approaches to pose determination. It can fully apply all the information used by past methods, and at the same time use new, weaker constraints on a possible match between image and model.

Our primary achievement in developing this approach has been a set of uniqueness results, analogous to the most basic uniqueness results for other approaches to pose determination. These results make precise the value of a loose correspondence between regions that is not based on specific local feature correspondences. At the same time, we also demonstrate that our basic approach applies to a wide variety of viewing transformations (similarity, affine, perspective), and to both 2-D and 3-D objects.

Finally, we have demonstrated the potential applicability of our method with experiments on real images. These show that we can correctly determine pose in spite of moderate amounts of occlusion, and normal sensor error. Our algorithm's performance on images with high perspective distortion also demonstrates the value of extending our method to perspective projection.

Acknowledgment

The authors wish to thank Ovadya Menadeva for his assistance in taking photographs and running the experiments.

References

[1] Alter, T. D. and D. Jacobs, 1994, "Error Propagation in Full 3D-from-2D Object Recognition", *CVPR*: 892-898.
 [2] Amenta, N., 1994, "Bounded boxes, Hausdorff distance, and a new proof of an interesting Helly-type theorem", *Proc. of*

the 10th Annual ACM Symposium on Computational Geometry: 340-347.

- [3] Ayache, N. and O. Faugeras, 1986, "HYPER: A New Approach for the Recognition and Positioning of Two-Dimensional Objects", *IEEE Trans. on PAMI*, **8**(1): 44-54.
 [4] Baird, H., *Model-Based Image Matching Using Location*, MIT Press, Cambridge, 1985.
 [5] Basri, R. and D. W. Jacobs, 1994. "Recognition using region correspondences", Forthcoming Technical Report.
 [6] Breuel, T., 1991, "Model Based Recognition using Pruned Correspondence Search," *CVPR*: 257-268.
 [7] Canny, J., 1983. "Finding edges and lines in images". *MIT A. I. Memo #720*.
 [8] Cass, T., 1992, "Polynomial Time Object Recognition in the Presence of Clutter, Occlusion and Uncertainty," *ECCV*: 834-842.
 [9] Conway, J.B., 1990. *A Course in Functional Analysis*. Springer-Verlag.
 [10] Fischler, M.A. and Bolles, R.C., 1981. "Random sample consensus: a paradigm for model fitting with application to image analysis and automated cartography". *Com. of the ACM* **24**(6): 381-395.
 [11] Forsyth, D., Mundy, J. L., Zisserman, A., Coelho, C., Heller, A., and Rothwell, C., 1991. "Invariant descriptors for 3-D object recognition and pose". *IEEE Trans. on PAMI*, **13**(10): 971-991.
 [12] Hu M.K., 1962. "Visual pattern recognition by moment invariants". *IRE Trans. on Info. Theory*, **IT-8**: 169-187.
 [13] Huttenlocher, D., G. Klanderman, and W. Rucklidge, 1993, "Comparing Images Using the Hausdorff Distance," *IEEE Trans. on PAMI*, **15**(9):850-863.
 [14] Huttenlocher, D., J. Noh, and W. Rucklidge, 1993, "Tracking Non-Rigid Objects in Complex Scenes," *ICCV*: 93-101.
 [15] Huttenlocher, D.P., and Ullman, S., 1990 "Recognizing Solid Objects by Alignment with an Image", *IJCV*, **5**(2): 195-212.
 [16] Jacobs, D., 1992, "Recognizing 3-D Objects Using 2-D Images". *MIT A.I. Memo 1416*.
 [17] Keren, D., J. Subrahmonia, G. Taubin, D. Cooper, 1992, "Bounded and Unbounded Implicit Polynomial Curves and Surfaces, Mahalanobis Distances, and Geometric Invariants, for Robust Object Recognition". *DARPA IUW*: 769-777.
 [18] Kriegman, D. and J. Ponce, 1990, "On Recognizing and Positioning Curved 3-D Objects from Image Contours," *IEEE Trans. on PAMI*, **12**(12): 1127-1137.
 [19] Lowe, D., 1985, *Perceptual Organization and Visual Recognition*, Kluwer Academic Publishers.
 [20] Pavlidis, T. and S. Horowitz, 1974, "Segmentation of Plane Curves," *IEEE Trans. on Computers*, **C**(23):860-870.
 [21] Richard C.W. and Hemami H., 1974. "Identification of three dimensional objects using Fourier descriptors of the boundary curve". *IEEE Trans. on Systems, Man and Cybernetics*, **4**(4): 371-378.
 [22] Ullman, S. and Basri, R., 1991. "Recognition by linear combinations of models". *IEEE Trans. on PAMI*, **13**(10): 992-1006.
 [23] Weiss, I., 1988. "Projective invariants of shape", *DARPA IUW*: 1125-1134.