International Journal of Computer Vision 25(2), 145–166 (1997)

Recognition Using Region Correspondences

RONEN BASRI*
Department of Applied Math., The Weizmann Inst. of Science, Rehovot, 76100, Israel
ronen@wisdom.weizmann.ac.il

DAVID W. JACOBS
NEC Research Institute, 4 Independence Way, Princeton, NJ 08540, USA
dwj@research.nj.nec.com

Received February 23, 1995; Accepted October 14, 1996

Abstract. Recognition systems attempt to recover information about the identity of observed objects and their location in the environment. A fundamental problem in recognition is *pose estimation*. This is the problem of using a correspondence between some portions of an object model and some portions of an image to determine whether the image contains an instance of the object, and, in case it does, to determine the transformation that relates the model to the image. The current approaches to this problem are divided into methods that use "global" properties of the object (e.g., centroid and moments of inertia) and methods that use "local" properties of the object (e.g., corners and line segments). Global properties are sensitive to occlusion and, specifically, to self occlusion. Local properties are difficult to locate reliably, and their matching involves intensive computation.

We present a novel method for recognition that uses region information. In our approach the model and the image are divided into regions. Given a match between subsets of 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. The new approach combines many of the advantages of the previous two approaches, while avoiding some of their pitfalls. Like the global methods, our approach makes use of region information that reflects the true shape of the object. But like local methods, our approach can handle occlusion.

Keywords: object recognition, occlusion, affine, perspective, regions, pose estimation, uniqueness, two-dimensional, three-dimensional, model

1. Introduction

One of the key problems of visual object recognition is to determine how best to effectively represent our knowledge of objects. In this paper we introduce a new representation in which the regions (i.e., parts) of objects are expressed directly as sets of points avoiding the need to compute more complex descriptions, (e.g., corners, or moments) which may be difficult to obtain in the presence of noise and occlusion. Although our representation is simple and flexible, it is not obvious that we can use it effectively for recognition. However, we demonstrate the utility of our representation by showing how to use it to solve one of the central problems of recognition, that of *pose estimation*. Moreover, our solution method is compatible with one based on local geometric features (points and lines), allowing us

*The research of Ronen Basri was supported in part by the Israeli Ministry of Science, Grant No. 6281. Ronen Basri is an incumbent of Arye Dissentshik Career Development Chair at the Weizmann Institute.

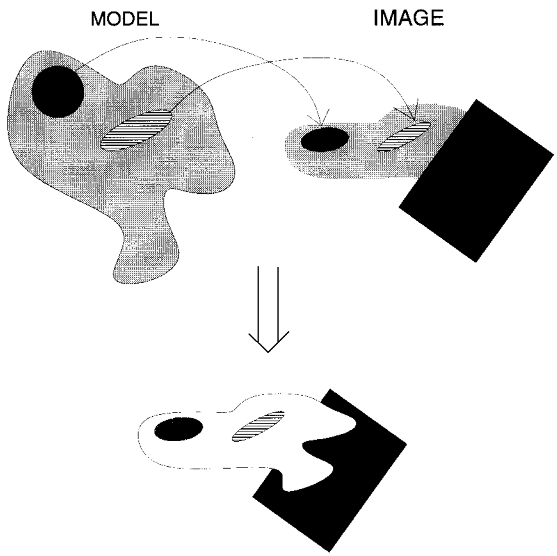

Figure 1. An example of region-based pose determination. The two matched regions determine the pose under an affine transformation.

to combine our representation with these, when they are also applicable.

Pose estimation is the problem of determining the transformation that relates the model to the image given a correspondence between some portions of an object model and some portions of an 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 Fig. 1. We show that in general a small number of region correspondences determine the correct pose of the object uniquely. We further analyze the degenerate cases and present experiments that demonstrate the usefulness of the method.

The novelty of our method lies in our use of a simple, direct representation based on region information. We assume that some part of the object model has been matched to a possibly occluded region in the image. This means that every point in the image region should be matched to some, as yet unknown point in the model part by the correct transformation. Previous methods typically attempt to solve this problem by computing and matching more distinctive features of the object and image, whereas our method will determine the correct transformation directly from a region correspondence.

For example, one set of previous methods represents objects or their parts using "global" features. In one example of a global method an object is represented in some canonical coordinate frame obtained by normalizing certain properties of the object (e.g., the origin is set at the object's center of mass, and the axes are aligned with its principal moments). Given an image, the region that contains the object is first segmented from the image. The corresponding properties of the object in the image are then computed and used to bring the object into the canonical description. Higher order moments, or other global descriptors may also be used to identify the object (Hu, 1962; Richard and Hemami, 1974; Dudani et al., 1977; Persoon and Fu, 1977; Sadjadi and Hall, 1980; Reeves et al., 1988). The advantage of this approach is that it is computationally efficient, since processing the image can be carried out independently of the model. The main difficulty with this approach is that it requires a good segmentation of the object, and it is sensitive to occlusion, and in particular to self occlusion. This makes the method unsuitable for recognizing 3-D objects from single 2-D images, or to recognition in cluttered scenes.

In a second previous approach, local features are extracted from the model and from the image. Subsets of model features are matched to subsets of image features, and this match is used to recover the alignment transformation. This has been done using point features (Fischler and Bolles, 1981; Horaud, 1987; Huttenlocher and Ullman, 1990; Ullman and Basri, 1991; Alter and Grimson, 1993; Alter and Jacobs, 1994), line segments (Lowe, 1985; Ayache and Faugeras, 1986; Rothwell et al., 1992), vertices (Thompson and Mundy, 1987), and distinguished points on curves such as inflection points or bitangents (Joshi et al., 1994, Rothwell et al., 1992). Other methods use algebraic descriptions of portions of contours to compute alignments (Kriegman and Ponce, 1990; Weiss, 1993; Forsyth et al., 1991; Subrahmonia et al., 1996). These can contain rich information about shape by describing larger portions of the object, although such descriptions may also be more sensitive to occlusion, and more difficult to compute reliably from an image.

By relying on local properties, these methods can be more robust than global ones. Typically we must isolate an entire shape to extract its global properties. However, we can often find local features using only fragments of contours of an object. This can make local methods resistant to partial occlusion or to segmentation failures.

One disadvantage of local methods is that they can be computationally expensive. For instance, for polyhedral objects all triplets of model points must be tried against all triplets of image points to guarantee that a solution is found ($O(m^3n^3)$ matches).

We feel that the relative merits of local and region-based methods will depend on which type of representation can best capture the shape of the object. In many situations, it has proven extremely difficult to reliably locate local features in images of 3-D objects. The intensities produced by a portion of a 3-D object will tend to vary significantly with changes in illumination and viewpoint. This leads to variations in the edges produced by the object, or in the performance of other intensity based approaches to local feature detection. Also, even when the contour of an object is accurately detected, it can be quite difficult to reliably find local features on contours that are even slightly curved. For this reason, most demonstrations of 3-D recognition systems have been limited to objects that are largely polyhedral. These problems can also occur in the recognition of planar objects, especially of curved planar objects.

We believe that these difficulties arise due to a fundamental difficulty with representations that use local features based on the contour of an object. As an example, in Fig. 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 general, 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.

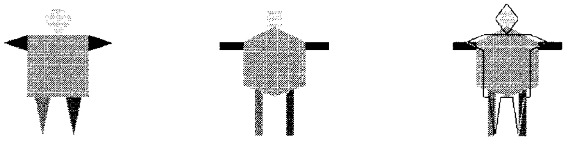

Figure 2. Two polygons that differ in features yet share the same overall structure. On the right, an alignment of the two figures using our method.

Because of these difficulties, we have developed an alternate representation based directly on the regions that make up objects, rather than on the contours that bound these regions. In our approach the model and the image are divided into convex regions. (Convexity is not a limitation, however, since concave entities can be replaced by their convex hulls.) Each region is described directly as a set of points, so that our representation captures all available shape information without extracting higher level descriptions. Given a match between model regions and image regions that may be partially occluded, 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. We will also show that the method can be extended to the recognition of a 3-D object in a 2-D image.

A number of previous recognition systems have also focused on recognizing objects in terms of their component parts (e.g., (Bajcsy and Solina, 1987; Biederman, 1985; Binford, 1971; Brooks, 1981; Hoffman and Richards, 1985; Marr and Nishihara, 1978; Pentland, 1987). Most of these approaches are concerned with determining the perceptual category of objects. They further assume that the part structure of the object is invariant under changes in viewing position and across objects in the same perceptual category. Often, complex descriptions of object parts are computed, based on the shape of the parts' bounding contours. Recognition is performed by comparing two graphs representing the descriptions and spatial relations between the parts of the object. In contrast, our method estimates the pose of specific objects whose exact shape is known. In addition, our method does not require the spatial relations between the parts to be invariant in different views. Also, the method can recognize objects when only a small subset of their parts (as few as two) are available, and it allows for the use of local features in addition to the parts. However, the primary difference between our work and previous approaches to parts-based recognition is the simple representation that we use, which does not rely on computing high-level, contour based representations of parts.

Our method relies on two different sets of possible constraints. We define these constraints precisely in Section 2. Intuitively, the first constraints require the transformation to project as much of each object region as possible so that it lies inside the corresponding image region. The second set of constraints requires that as much of each image region as possible is covered by the projection of each model region. In the case of a 2-D object and a 2-D image, we are able to present a very efficient algorithm that enforces either set of constraint. The second set is especially useful when the object is partially occluded, since it requires the visible image regions to be explained by the object, but does not require the object regions to be fully explained by the image. The algorithm is designed to find poses that will match entire regions, without requiring an explicit correspondence between a local feature in the region. Thus a pose is found that matches portions of the model and image well without requiring us to isolate features in these regions, or to hypothesize a correspondence between specific features. However, if a correspondence between model and image point features or line features should be available, this local feature information can be fully used by the algorithm as well. Because our algorithm is implemented by reducing our problem to a linear program, it can be run quite efficiently.

We describe the algorithm and its behavior more thoroughly in Section 3. We then address a number of questions concerning the algorithm's performance. First, in the error-free case, we determine exactly when sufficient information exists to uniquely determine the pose of an object. These results are analogous to results for local features showing, for example, the minimal number of point or line correspondences needed to determine an object's pose. We show that in the general case of planar models, a correspondence between two convex regions suffices to uniquely determine pose (or three regions, when we allow for perspective effects). Next, in Section 4 we show how to extend these ideas to 3-D objects as well. We show that in this domain we can solve the first set of constraints using linear programming. And in the case of 3-D objects that have planar regions (that are not necessarily coplanar), we again show the minimum number of region correspondences needed to solve uniquely for the object pose. This demonstrates that to a significant extent, our algorithms can be applied to the recognition of 3-D objects in 2-D images. Finally, in Section 5 we show experiments on a number of real objects to demonstrate the system's performance. An abbreviated description of these results has previously appeared in (Basri and Jacobs, 1995).

In summary, our new approach has several practical advantages over previous approaches.

- The new approach does not require an exact localization of features. Good estimation of the alignment transformation can be obtained even when the boundaries of the regions can only roughly be localized. This makes the method particularly suitable as a second stage for color and texture classifiers.
- The method is computationally efficient. This is due to two reasons. (1) Unlike points and lines, regions can often be identified by their properties (e.g., color and texture). (2) In certain cases a match between two model and image regions is already sufficient to recover the alignment transformation. This reduces significantly the combinatorics of the matching process.
- The method handles objects with smooth curved surfaces. Predicting the position of the contours on the object is not required for estimating the alignment transformation.
- The method can handle certain kinds of occlusion. When the model is planar, we can make use of image regions that are an arbitrary, occluded subset of the model region. For 3-D models we can handle self-occlusion, and also we can make use of image regions that are partially occluded, when the occluding lines have been identified.

These advantages indicate that the proposed method has the potential to recognize objects in domains that have proven difficult for existing systems to handle. In particular, the system offers the potential to reliably recognize curved 3-D objects, on which local features and contours may be difficult to find. Also, by focusing on region information, our algorithm has potential to offer a better way of comparing the shape of different objects that are instances of the same class of objects.

2. Problem Definition

Below we consider the following problem. Given a set of model regions $V_1, \ldots, V_k \subset \mathcal{R}^d$, $d \in \{2, 3\}$, and a corresponding set of image regions $R_1, \ldots, R_k \subset \mathcal{R}^2$ determine the transformation $T \in \mathcal{T}$ that maps every model region V_i to its corresponding image region R_i $(1 \leq i \leq k)$.

Throughout the paper we assume that the 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.

We consider a few variants of this problem. The model may be 2-D or 3-D. If 2-D, the allowed transformations, $\mathcal{T}$, are either the similarity (rotation, translation, and uniform scaling), affine (linear and translation), or projective transformations. If 3-D, we consider affine transformations followed by either an orthographic or perspective projection.

It will be very useful to distinguish between two sets of constraints on the transformation.

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 model point $\vec{p} \in V_i$ (that is, $TV_i \supseteq R_i$).

The backward constraints are particularly useful, since they capture fully our knowledge of pose when we allow for the image regions to be arbitrarily occluded.

The problem of determining the transformation that perfectly aligns a set of model regions to their corresponding image regions can be described as finding a transformation that satisfies both the forward and the backward constraints. This 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.

Below we consider the problem of computing the transformation T, of a family of transformations $\mathcal{T}$, that is consistent with either just the forward constraints or just 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.

3. The 2-D Problem

In this section we analyze the case of matching 2-D model regions to 2-D image regions. We consider three sets of allowed transformations, similarity, affine, and projective. A similarity transformation is composed of a rigid transformation and uniform scaling. An affine transformation may include, in addition, stretch and shear. 2-D affine transformations represent the set of transformations relating two scaled-orthographic images of a plane. Projective transformations relate two perspective images of a planar object.

We begin by defining the one-way constraints (either forward or backward). We show that in all three cases (similarity, affine, and projective) 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.

Our method is based on Baird's (1985) insight that matching features to convex regions leads to linear constraints on the set of possible transformations relating the two. Cass (1992) used this insight to produce the first polynomial-time algorithm guaranteed to match the maximum number of consistent features. Breuel (1991) also uses these insights to make potentially exponential constraint-based search into a worst-case polynomial time algorithm. However, our work differs significantly from these approaches, which focus on matching simple, local features, and do not make use of region-based information.

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. There are two possible sets of one-way constraints, the forward constraints (where model regions are required to project entirely inside their corresponding image regions) and the backward constraints (where image regions are required to lie entirely inside the projection of the their corresponding model regions).

We formulate the forward constraints as follows. Given a convex model region 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

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

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

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

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 regions $V_1, \ldots, V_k$ and corresponding image regions $R_1, \ldots, R_k$ 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 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

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

Note that given the model region V and its corresponding image region R the model point $\vec{p}$ is known (it may be any of the points in V), and so is the constraint line $Au + Bv + C \geq 0$ (which may be any of the tangent lines to R). 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

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

The forward constraint in this case is given by

$$A(ax + by + c) + B(-bx + ay + d) + C \geq 0. \tag{5}$$

This constraint is linear in the transformation parameters. (This parameterization is used and explained, for example, in Baird (1985). Baird pointed out that a linear bound on the location of a transformed model point leads to a linear constraint on the feasible transformations.) Denote

$$\vec{w}^T = (a, b, c, d),$$

and

$$\vec{g}^T = (Ax + By, Ay - Bx, A, B),$$

we can rewrite the forward constraint (4) as

$$\vec{g}^T \vec{w} \geq -C. \tag{6}$$

When T is restricted to be a Euclidean transformation (with no scaling) an additional non-linear constraint is obtained:

$$a^2 + b^2 = 1. \tag{7}$$

This constraint is not used in our scheme.

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

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix}. \tag{8}$$

We can now write the forward constraint for the affine case in the following form

$$A(ax + by + e) + B(cx + dy + f) + C \geq 0. \tag{9}$$

This constraint is linear in the transformation parameters. Denote

$$\vec{w}^T = (a, b, c, d, e, f)$$

the vector of unknown transformation parameters, and

$$\vec{g}^T = (Ax, Ay, Bx, By, A, B)$$

we can rewrite the forward constraint (8) as

$$\vec{g}^T \vec{w} \geq -C \tag{10}$$

Projective Transformation. In order to extend our formulation of the forward constraints to handle projective transformations we should first overcome one inherent difficulty. Our formulation relies on matching convex model regions to convex image regions.

Projective transformations, however, may transform a convex region to a non-convex one. This difficulty is circumvented by noticing that under projectivity convex shapes are mapped to non-convex ones only when the object crosses the image plane (or, in other words, when the vanishing line intersects the object). Since under perspective projection the image plane always lies between the object and the focal point it is guaranteed that convex regions on the object will produce convex regions in the image. The subset of projective transformations relevant to recognition therefore preserves convexity.

We now show how to formulate the one-way constraints in the projective case. T can be expressed in the form

$$\alpha \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} a & b & e \\ c & d & f \\ g & h & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{11}$$

for some arbitrary scalar factor α. (Note that, WLOG, we set T_{33} to be 1.) Thus,

$$u = \frac{ax + by + e}{gx + hy + 1} \qquad v = \frac{cx + dy + f}{gx + hy + 1}. \tag{12}$$

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) we obtain

$$\begin{aligned} A(ax + by + e) + B(cx + dy + f) \\ + C(gx + hy + 1) \geq 0 \end{aligned} \tag{13}$$

Again, this constraint is linear in the transformation parameters. Denote

$$\vec{w}^T = (a, b, c, d, e, f, g, h)$$

the vector of unknown transformation parameters, and

$$\vec{g}^T = (Ax, Ay, Bx, By, A, B, Cx, Cy)$$

we can rewrite the forward constraint (12) as

$$\vec{g}^T \vec{w} \geq -C \tag{14}$$

Backward Constraints. In the 2-D case models and images are interchangeable, and so the backward constraints can be defined in the same way as the forward constraints. Again, for affine, similarity, and projective transformations the constraints are linear and they take the form

$$\vec{g}^T \vec{w} \geq -C, \tag{15}$$

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 regions (after the model and the image are brought into alignment), but the inclusion may be strict due to the occlusion.

3.2. Solving a System of 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 regions 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 regions the number of independent constraints is finite. The system of constraints in this case is completely defined by the vertices of the model regions 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 (Basri and Jacobs, 1994).

Given a finite set of constraints

$$\vec{g}_i^T \vec{w} \geq c_i, \qquad i = 1, \ldots, n \tag{16}$$

we seek a vector of parameters $\vec{w}$ that is consistent with the constraints. Denote by G the matrix of rows $\vec{g}_i$, and by $\vec{c}$ the vector of c_i's. We may write

$$G\vec{w} \geq \vec{c} \tag{17}$$

where the $\geq$ sign applies separately to each of the components.

Solving the one-way problem (17) involves finding a linear discriminant function. One method of finding a linear discriminant is by using linear programming.

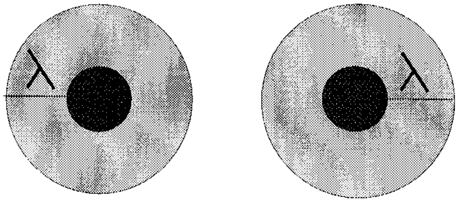

Figure 3. The dark circles are positioned by the similarity transformation that maximizes λ relative to the larger, shaded circles.

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, in the following way.

$$\begin{aligned} &\max \lambda \\ s.t. \quad & G\vec{w} \geq \vec{c} + \lambda\vec{1} \end{aligned} \tag{18}$$

A solution to (17) exists if and only if a solution to (18) with $\lambda \geq 0$ exists. (Note that other objective functions, e.g., the perceptron function, can be used for recovering $\vec{w}$, see e.g., (Duda and Hart, 1973) for a discussion of solutions to the linear discriminant functions problem.)

When $\lambda \geq 0$ its value represents the minimal distance of a point to any line bounding the region (Fig. 3). Maximizing λ amounts to attempting to contract the model region 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. Therefore, when sensing error prevents us from fully satisfying the constraints we will find an approximate solution. ($|\lambda|$ in this case represents a maximum norm, and so it is related to the Hausdorff metric. For work on Hausdorff matching, see (Huttenlocher et al., 1993a, 1993b). Also, Amenta (1994) specifically discusses the efficient Hausdorff matching of convex shapes undergoing translation and scaling.)

Solving the system (18) may result in over-contraction. Consider, for example, the case of matching a single model region V to a single image region R. The forward constraints restrict the set of possible transformations to those that map every point $\vec{p} \in V$ inside the region R. Assume T is a feasible transformation, that is $TV \subseteq R$, then applying any contracting factor $0 \leq s \leq 1$ to V would also generate a valid solution; namely, $T(sV) \subseteq R$. (We assume here without the loss of generality that the origin of the model and the image are set at the centroid of V and R respectively.) Consequently, the case of matching one region necessarily introduces multiple solutions. The solution picked by Eq. (18) is the one with $s = 0$. This will contract V to a point, which is then translated to the point inside R furthest from any of its bounding tangent lines. This solution produces the largest value of λ. Clearly, the case of matching one region cannot be solved by the forward constraints alone.

In what follows we prove that generally if the model contains two or more non-overlapping regions the solution is unique. We specify the degenerate cases and show that they can be predicted from the model alone.

3.3. Uniqueness Theorems

In this section we establish the conditions under which a one-way region matching problem has a unique solution. We state the problem as follows:

Problem Statement. We are given a set of convex model regions and corresponding convex image regions. The image regions are produced by applying either a similarity, affine or projective transformation to the model regions. If we consider the similarity (affine, projective) transformations that will project the model regions entirely inside the corresponding image regions, under what circumstances is the transformation that does this uniquely determined? Note that clearly, 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 when we are considering invertible transformations.

We begin this section by proving a basic lemma (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 transformation (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 regions entirely inside themselves.

Using Lemma 1 we show that in the similarity case two distinct (non-intersecting) model regions and their corresponding image regions determine the transformation uniquely. In both the affine and the projective cases we show that three regions positioned such that no straight line passes through all three regions determine the transformation uniquely. Then, we derive necessary and sufficient conditions for two regions to determine a unique solution. In the affine case these conditions imply that for most pairs of regions the transformation is determined uniquely. Degenerate cases are analyzed in Section 3.4. In the projective case, however, we show in (Basri and Jacobs, 1994) that for example, due to these conditions, the transformation is never determined uniquely for any pair of triangles or ellipses. The analysis of the three transformation groups appears later in this section. Section 3.3 discusses the conditions for uniqueness in the similarity case. Section 3.3 discusses the conditions for uniqueness in the affine case, and Section 4 discusses the conditions for uniqueness in the projective case. We conclude (Section 5) with a discussion of the uniqueness of the one-way matching problem when points and line segments are used as regions.

We now turn to showing that uniqueness is dependent on the model alone.

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

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

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

Theorem 1. *Let $V_1, V_2 \subseteq \mathcal{R}^2$ be two distinct convex closed regions ($V_1 \cap V_2 = \emptyset$). Then, the solution to the one-way matching problem with these regions 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 regions to inside themselves then, by Brouwer's fixed point theorem (Conway, 1990), 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,

$$T(\vec{p}_i) = \vec{p}_i \qquad i = 1, 2.$$

(Note that $\vec{p}_1 \ne \vec{p}_2$ 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 regions to within themselves, and so T must be the identity transformation. □

Notice that Theorem 1 requires that the two regions will be completely disjoint. If the two regions intersect in a curve, or share even only a single point, then it will always be possible to contract the model to a single point and map it to the common intersection. In contrast, if the two regions are disjoint the solution to the one way constraints will be unique even if the regions contain symmetries or isomorphisms of any kind.

Affine Transformations. In this section we handle the affine case. We first show that an affine transformation is uniquely determined by three regions that cannot be traversed by any straight line. Later we derive a necessary and sufficient condition for two regions to determine a unique solution.

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

Proof: Similar to Theorem 1, assume T is an affine transformation that maps the regions to inside themselves. Then there exist three points that are fixed with respect to T. Since no straight line pass through all three regions the three fixed points are non-collinear, and so they determine the identity as the only affine transformation that maps the regions to inside themselves. Therefore $T = I$. □

We now turn to showing that the number of regions required to determine the affine transformation uniquely is in general two. Theorem 4 below establishes that two distinct regions determine the transformation uniquely unless the regions can be contracted such that both regions shrink entirely inside themselves. This property is used further in Section 3.4 to characterize the degenerate cases.

Theorem 3. *Let $V_1, V_2 \subseteq \mathcal{R}^2$ be two distinct closed regions. Then, the solution to the one-way matching problem with these regions 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*

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

(*see Fig.* 4).

Proof: One direction is straightforward. Assume $T_{l,\vec{v}}$ contracts the regions 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:

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

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 $T(V_i) \subseteq V_i$ $(i = 1, 2)$. We next show that T is $T_{l,\vec{v}}$. Since T maps the two regions 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,

$$T(\vec{p}_i) = \vec{p}_i \qquad i = 1, 2.$$

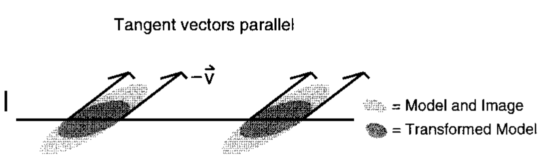

Figure 4. Two model regions lead to non-unique affine transformations when a line l, exists such that the tangents at all intersection points are parallel. In this case, the regions can contract towards l in the direction $\vec{v}$.

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:

$$T = \begin{pmatrix} 1 & a \\ 0 & b \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

(So that every point $(x, 0)^T$ is mapped to itself.) Denote the angle between $\vec{v}$ and l by ψ, then contraction in a direction $\vec{v}$ toward l is expressed by

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

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. In general, such a line will not exist. An analysis of the degenerate cases is given in Section 3.4.

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 regions such that no straight line passes through any three of the regions determine the projective transformation uniquely. (Simply, the four regions induce four fixed points, and four points such that no three are collinear determine the projective transformation.) Four, however, is not the minimal number of regions that determine a unique solution. We are able to show:

Theorem 4. *Let $V_1, V_2, V_3 \subseteq \mathcal{R}^2$ be three closed regions with non-zero areas such that there exists no straight line passing through all three regions. Then, the solution to the one-way matching problem with these regions as a model under a projective transformation is unique.*

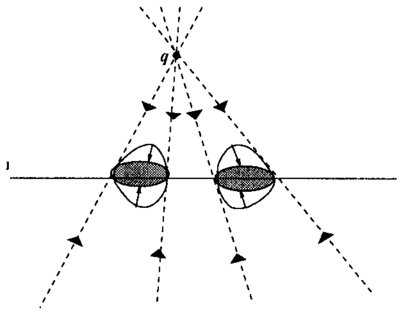

Figure 5. Two model regions lead to non-unique projective transformations when a line l, exists such that the tangents at all intersection points meet at a single point q. In this case, the regions can contract towards l in the directions emanating from q.

And, we can show:

Theorem 5. *Let $V_1, V_2 \subseteq \mathcal{R}^2$ be two distinct closed regions with non-zero areas. Then, the solution to the one-way matching problem with these regions as a model under a projective transformation is non-unique 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*

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

(*see Fig.* 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 (Basri and Jacobs, 1994).

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 below.

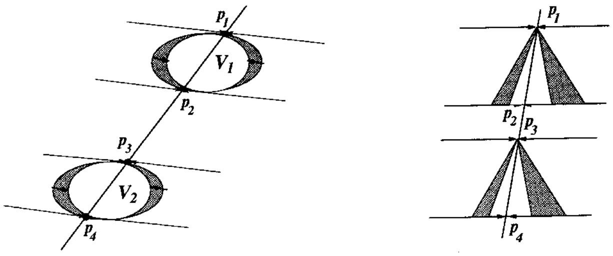

Figure 6. Cases when solution is non-unique: when there exists four collinear points on the boundaries of the two regions with parallel tangents (left), or when either of these points is a vertex and the line connecting the four points pierces the region (right).

Theorem 6. *Using just the one-way constraints*:

- *A similarity transformation is determined uniquely from two points, from two line segments, or from a combination of a point and a line segment.*
- *An affine transformation is determined uniquely from any three points or line segments, provided that no line intersects all three features.*
- *An affine transformation is determined uniquely from two or more line segments and any number of points even when a line does intersect all features, provided that there are at least two line segments not contained in the intersecting line that are parallel.*
- *An affine transformation is not uniquely determined when all points and line segments lie on a single line, or when line segments that do not lie on this line are parallel.*
- *A projective transformation is determined uniquely from four points such that no three are collinear, and by three line segments in general position, that are not all intersected by a single line.*

The proof of this theorem follows directly from the proofs of previous theorems.

An important advantage of the proposed formulation is that it can handle combinations of feature points, line segments, and regions in the same framework.

3.4. Degeneracies

In the previous section we showed that in general two distinct regions determine the alignment transformation uniquely. No degenerate cases exist if the alignment transformation is restricted to be a similarity transformation. The affine case, however, introduces degeneracies, and a third region may be required to disambiguate a solution. In this section we analyze the conditions for the occurance of degeneracies. We introduce necessary and sufficient conditions for the existence of degeneracies and complete the analysis with several examples.

Suppose that $V_1, V_2 \subseteq \mathcal{R}^2$ are two distinct regions ($V_1 \cap V_2 = \emptyset$). Let l denote a line passing through both regions, and let $\vec{v}$ denote a direction, different from the direction of l. Denote the entry (or exit) points of l into V_1 by $\vec{p}_1, \vec{p}_2$ and into V_2 by $\vec{p}_3, \vec{p}_4$. Using Theorem 4 it can be shown that this regions are degenerate under an affine transformation if and only if there exist l and $\vec{v}$, such that each point $\vec{p}_j$ satisfies either one of the two conditions:

1. The tangent to the boundary of V_i at p_j is parallel to $\vec{v}$, or
2. p_j is a vertex, l intersects the interior of V_i, and the line through p_j with direction $\vec{v}$ does not intersect the inside of V_i.

These conditions provides a complete characterization of the degenerate cases, and so we can use them to analyze any given model. Below we analyze the cases of objects composed of smooth bounded regions and objects composed of polygons.

Suppose that both V_1 and V_2 have smooth boundaries. Then, due to condition 1, V_1 and V_2 are degenerate under an affine transformation if and only if there exist four collinear points on the boundaries of the regions with parallel tangents. Consequently, two circles are always degenerate, since the line connecting their centers penetrates the circles at points with parallel tangents. (The tangents at these points are perpendicular to l, see Fig. 7.) In contrast, two ellipses in general position are not degenerate.

Suppose that both V_1 and V_2 are two distinct polygons. Then, the two conditions take the following form: V_1 and V_2 are degenerate if and only if there exist four collinear points on the boundaries $\vec{p}_1, \vec{p}_2 \in V_1$ and $\vec{p}_3, \vec{p}_4 \in V_2$ which satisfy the following condition. Let l be the line through $\vec{p}_1, \ldots, \vec{p}_4$. For every $1 \leq j \leq 4$, denote the angle between l and the boundary of the shape at p_j by $\alpha_{1j} \leq \alpha_{2j}$ as in Fig. 8 ($\alpha_{1j} < \alpha_{2j}$ if p_j is a vertex and $\alpha_{1j} = \alpha_{2j}$ otherwise). Then degeneracy occurs if $\bigcap_{j=1}^{4} [\alpha_{1j}, \alpha_{2j}] \neq \emptyset$.

As an example we analyze below the case of a model consisting of two distinct triangles. Let l denote the fixed line and $\vec{v}$ denote the direction of contraction. For each triangle the theorem restricts l and $\vec{v}$ to the two following cases:

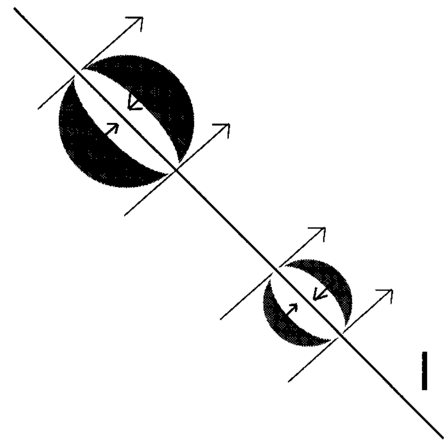

Figure 7. Two circles always lead to degenerate solutions under an affine transformation, as shown in the figure. A line through their centers intersects them at points with parallel tangents, allowing contraction in the direction perpendicular to this line.

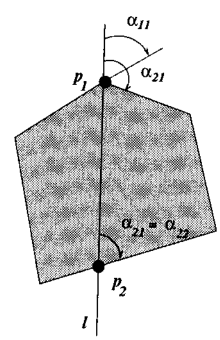

Figure 8. Notation: when l intersects the boundary of a polygon at a vertex (as in $\vec{p}_1$) α_{11} and α_{21} are the angles between the two sides emanating from $\vec{p}_1$ and the positive direction of l. When l intersects the boundary of a polygon at a side (as in $\vec{p}_2$) $\alpha_{21} = \alpha_{22}$ is the angle between the side and the positive direction of l.

1. l could be a side of a triangle, and $\vec{v}$ could be between α_1 and α_2 (where α_1 and α_2 denote the angles between the positive direction of l and each of the two other sides of the triangle), or
2. l could go through a vertex and a side of the triangle, and $\vec{v}$ could be in the direction of the side.

From this analysis we obtain that a pair of distinct triangles are degenerate only in the following three cases (see Fig. 9):

1. If the triangles contain four collinear vertices. Denote by l the straight line through the four vertices, by α_1, α_2, β_1, and β_2 the angles between the sides

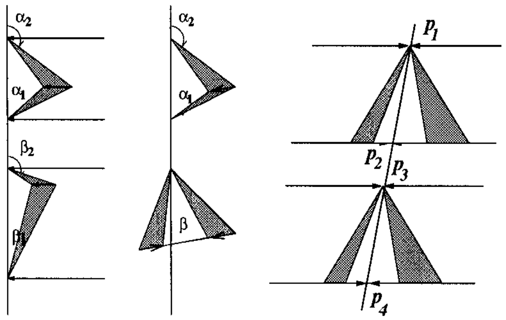

Figure 9. Degenerate sets of distinctive triangle pairs. Two triangles with four collinear vertices (left), with three collinear vertices (middle) and parallel sides (right). The conditions that makes these pairs of triangles degenerate are specified in the text.

of the triangles and the positive direction of l, and by θ the angle between $\vec{v}$, the direction of contraction, and l. Since for contraction to be possible the lines parallel to $\vec{v}$ through the four vertices must not pierce the inside of the triangles we obtain that $\alpha_1 \leq \theta \leq \alpha_2$ and $\beta_1 \leq \theta \leq \beta_2$. Contraction in this case is therefore possible if the ranges $[\alpha_1, \alpha_2] \cap [\beta_1, \beta_2] \neq \emptyset$.

2. If three of the vertices are collinear, denote by l the line connecting the three vertices, l must pierce one of the triangles in its side. This side determines the direction of contraction. Denote by β the angle between this side and the positive direction of l, and, as before, denote by α_1 and α_2 the angles between the positive side of l and the sides of the other triangle, now contraction is possible if $\beta \in [\alpha_1, \alpha_2]$.
3. Contraction is possible also if two sides of the triangles are parallel and the line connecting the two opposing vertices goes through the two triangles. In this case l is the line connecting the vertices and $\vec{v}$ is the direction of the parallel sides.

These are the only cases of degenerate triangles. N-sided polygons produce essentially the same results. The only difference is that a many-sided polygon can also have two parallel sides, which leads to one more type of case.

3.5. *Summary*

We have shown how to precisely formulate the one-way constraints, and how to efficiently find transformations that satisfy these constraints. The backward constraints express our state of knowledge exactly when we have matched parts of an object without any more precise correspondence between features of these parts, and when we allow for arbitrary amounts of occlusion. The uniqueness results are necessary to show that such a weak, region-based correspondence can be sufficient to determine pose. Note that although these results are demonstrated for the occlusion-free case, they are also significant when there is occlusion. Suppose, for example, that we match five model parts to five image regions, and solve for an affine transformation that satisfies the backward constraints. Our uniqueness results tell us that, in general, if two of the image regions are unoccluded, and the other three contain arbitrary amounts of occlusion, we will obtain the correct model pose. This is because occluded regions still contribute only more correct backward constraints, and cannot undermine the correctness of our pose. Therefore, we have a solution method that can tolerate large amounts of occlusion, and which does not need to know which portions of the region boundary are due to occlusion, and which parts represent the boundaries of the object.

4. The 3-D Problem

In this section we extend the method to matching 3-D model to 2-D image regions. This time we only consider the set of affine transformations in 3-D followed by either an orthographic or perspective projection. In the formulation below we match 3-D model volumes to 2-D image regions. We later introduce uniqueness results which apply only to planar model regions that are not necessarily coplanar. We begin by defining the one-way constraints. Unlike in 2-D, we consider only the forward constraints, since the back constraints cannot be expressed linearly. We then analyze the solution involving the application of the one-way constraints. The solution is again obtained by applying a linear program.

4.1. *One-Way Constraints*

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

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

$$AT_u(\vec{p}) + BT_v(\vec{p}) + C \geq 0 \tag{19}$$

The unknowns are the parameters of the transformation, T.

Affine + Orthographic Projection. First we consider a projection model consisting of a 3-D affine transformation followed by an orthographic projection. We will call this the orthographic case, to distinguish it from a 3-D affine transformation followed by perspective projection. Denote the linear part of T by R, where R is a non-singular 3×3 matrix with elements t_{ij}, and the translation part by $\vec{t} = (t_x, t_y, t_z)$. Then:

$$\begin{aligned} u &= t_{11}x + t_{12}y + t_{13}z + t_x \\ v &= t_{21}x + t_{22}y + t_{23}z + t_y. \end{aligned} \tag{20}$$

This projection model and its equivalent has been recently used by a number of researchers (Lamdan and Wolfson, 1988; Ullman and Basri, 1991; Koenderink and van Doorn, 1991; Tomasi and Kanade, 1992; Jacobs, 1992). It is also equivalent to applying scaled orthographic projection followed by a 2-D affine transformation (Jacobs, 1992), that is, taking a picture of a picture. Alternately, it is equivalent to a paraperspective projection followed by translation (Basri, 1996), where paraperspective is a first-order approximation to perspective projection (Poelman and Kanade, 1994; Sugimoto, 1996).

The forward constraint for the orthographic case becomes

$$\begin{aligned} &A(t_{11}x + t_{12}y + t_{13}z + t_x) \\ &\quad + B(t_{21}x + t_{22}y + t_{23}z + t_y) + C \geq 0. \end{aligned} \tag{21}$$

This constraint is linear in the transformation parameters. Denote

$$\vec{w}^T = (t_{11}, t_{12}, t_{13}, t_x, t_{21}, t_{22}, t_{23}, t_y)$$

the vector of unknown transformation parameters, and

$$\vec{g}^T = (Ax, Ay, Az, A, Bx, By, Bz, B)$$

we can again rewrite the forward constraints as

$$\vec{g}^T \vec{w} \geq -C. \tag{22}$$

Affine + Perspective Projection. Consider now the case of perspective projection. In this case

$$\begin{aligned} u &= \frac{f(t_{11}x + t_{12}y + t_{13}z + t_x)}{t_{31}x + t_{32}y + t_{33}z + t_z} \\ v &= \frac{f(t_{21}x + t_{22}y + t_{23}z + t_y)}{t_{31}x + t_{32}y + t_{33}z + t_z} \end{aligned} \tag{23}$$

where f is the focal length. The forward constraint $Au + Bv + C \geq 0$ implies that

$$\begin{aligned} &A\frac{f(t_{11}x + t_{12}y + t_{13}z + t_x)}{t_{31}x + t_{32}y + t_{33}z + t_z} \\ &\quad + B\frac{f(t_{21}x + t_{22}y + t_{23}z + t_y)}{t_{31}x + t_{32}y + t_{33}z + t_z} \\ &\quad + C \geq 0 \end{aligned} \tag{24}$$

Since we generally require the object to appear in front of the camera the term $t_{31}x + t_{32}y + t_{33}z + t_z$ must be positive. Thus, we obtain:

$$\begin{aligned} &Af(t_{11}x + t_{12}y + t_{13}z + t_x) \\ &\quad + Bf(t_{21}x + t_{22}y + t_{23}z + t_y) \\ &\quad + C(t_{31}x + t_{32}y + t_{33}z + t_z) \geq 0 \end{aligned} \tag{25}$$

Let

$$\vec{w}^T = (t_{11}, t_{12}, t_{13}, t_x, t_{21}, t_{22}, t_{23}, t_y, t_{31}, t_{32}, t_{33}, t_z)$$

contain the unknown transformation parameters, and let

$$\begin{aligned} \vec{g}^T = (&Afx, Afy, Afz, Af, Bfx, Bfy, Bfz, Bf, \\ &Cx, Cy, Cz, C) \end{aligned}$$

contain the known positional parameters, we obtain that

$$\vec{g}^T \vec{w} \geq 0 \tag{26}$$

In this case we obtain a homogeneous inequality and so solutions can be obtained only up to a scale factor.

To consider the rigid case for either orthographic or perspective projection, we have to add non-linear constraints enforcing the orthonormality of the row vectors of the rotation matrix, R. In the discussion below we confine ourselves to affine transformations under orthographic projection.

Backward Constraints. For the problem of matching 3-D models to 2-D images the back constraints cannot be specified in a straightforward way since the depth component of points in the image is eliminated by the projection. Consequently, in 3-D, only the forward constraints generate a set of linear constraints. The discussion below is restricted to the forward constraints.

Since we can only enforce the forward constraints in 3-D recognition, it is important to ask under what circumstances the forward constraints alone are valuable. First, our regions may consist of planar or curved 2-D portions on the "skin" of an object, such as surface markings or facets of an object. Such 2-D regions may frequently project without self-occlusion. Second, although a 3-D volume will always project with self-occlusion, this self-occlusion does not invalidate the forward constraints, since we do want a projection that takes all volume points inside the corresponding region. Third, the forward constraints may be used when there is known occlusion in a region. If we can identify the boundary of a region as due to an occlusion, we can eliminate it from the boundary, construct a region that is the maximal convex set of points known to belong to the region. The forward constraints will be correct when applied to such regions.

4.2. *Uniqueness Theorems*

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. As is explained in Section 3.2, the solution requires finding a linear discriminant function, and this can be done, in particular, by solving a linear program. In this section we will consider under what circumstances enforcing the forward constraints will produce a unique pose, when matching a 3-D model and a 2-D image under orthographic projection. The case of fully 3-D volumes is relatively challenging to analyze because such volumes always project to images with some self-occlusion. So we shall confine ourselves to the simpler case of a model that consists of planar regions that need not be mutually coplanar. For such a model we will show that the transformation is determined uniquely from the forward constraints when the model consists of four regions in general position. We will derive a sufficient condition for uniqueness when the model consists of three regions.

We are given a set of regions $V_1, \ldots, V_k \subset \mathcal{R}^3$ such that each V_i lies inside a plane P_i ($1 \le i \le k$) and a set of corresponding image regions $R_1, \ldots, R_k \subset \mathcal{R}^2$. We will assume without loss of generality that the image plane is $z = 0$. Denote the projection operator by Π. That is, Π transforms a 3-D object into a 2-D object by setting its z component to 0. In matrix form,

$$\Pi = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

By saying that the regions correspond, we mean that there exists some 3-D affine transformation, T, such that, for all i, $\Pi T V_i = R_i$. In matrix form, given a model point $\vec{p} = (x, y, z)$, we write:

$$T\vec{p} = \begin{pmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix}$$

We label the rows of T, $\vec{t}_i$. We wish to know under what circumstances T is unique.

Obviously, the effect that T has on the model's z component cannot be uniquely recovered. That is, $\vec{t}_3$ and t_z are never uniquely determined. We will say that two 3-D affine transformations are equivalent if they differ only in their third row and z translation. We would like to know if T is uniquely determined up to this equivalence relation.

Let us assume that there exists some affine transformation T' such that $\Pi T' V_i \subseteq R_i$. We wish to discover when T' must be equivalent to T. Define $T'' = T^{-1} T'$.

Lemma 2. *T is uniquely determined for the set of regions V_i if and only if it is uniquely determined for the set of regions QV_i, where Q is any 3-D affine transformation.*

Proof: Clearly, if we show this assertion in one direction, it must be true in the other, since the affine transformations form a group. Suppose T is not uniquely determined, i.e., that there exists T' not equivalent to T such that $\Pi T' V_i \subseteq R_i$. Then, let $W = TQ^{-1}$, $W' = T'Q^{-1}$. Clearly, W maps the regions QV_i to the regions R_i ($\Pi W Q V_i = R_i$), and W' maps the regions QV_i within these regions ($\Pi W' Q V_i \subseteq R_i$). We must show that W and W' are not equivalent.

To see this, we suppose that W and W' are equivalent, and show that this implies that T and T' are equivalent. First, abbreviate the rows of the linear parts of T and T' as: t_i, t'_i, abbreviate the columns of the linear part of Q^{-1} as q_i, and denote the translation of Q^{-1} by

$t_q = (q_x, q_y, q_z)$. The linear parts of W and W' are given by:

$$\begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix} (q_1 \quad q_2 \quad q_3), \begin{pmatrix} t_1' \\ t_2' \\ t_3' \end{pmatrix} (q_1 \quad q_2 \quad q_3)$$

Therefore, W equivalent to W' implies:

$$\begin{aligned} t_1 q_1 &= t_1' q_1 \\ t_1 q_2 &= t_1' q_2 \\ t_1 q_3 &= t_1' q_3 \\ t_2 q_1 &= t_2' q_1 \\ t_2 q_2 &= t_2' q_2 \\ t_2 q_3 &= t_2' q_3 \end{aligned}$$

This implies that $(t_1 - t_1')$ is orthogonal to q_1, q_2 and q_3. Since we assume that Q is non-singular, $t_1 = t_1'$. Similarly, $t_2 = t_2'$. The translation components of W and W' are given by:

$$\begin{pmatrix} t_1 t_q + t_x \\ t_2 t_q + t_y \\ t_3 t_q + t_z \end{pmatrix}, \begin{pmatrix} t_1' t_q + t_x' \\ t_2' t_q + t_y' \\ t_3' t_q + t_z' \end{pmatrix}$$

Therefore, $t_1 = t_1'$, $t_2 = t_2'$, and the equivalence of W and W' implies further that: $t_x = t_x'$, $t_y = t_y'$, and so T is equivalent to T', contradicting our assumption. □

Note that while similar to Lemma 1, Lemma 2 is much more limited. Lemma 1 says that in the planar case, uniqueness is independent of the image. Lemma 2 only states that in the 3-D case uniqueness is independent of the model's affine frame of reference. As we will discuss, in the 3-D case uniqueness is not independent of the image.

Lemma 3. *Given, as usual, model regions V_i, image regions R_i, and two affine transformations T and T' such that $\Pi T V_i = R_i$ and $\Pi T' V_i \subseteq R_i$, there exists a point $\vec{p}_i$ in each region V_i such that $\Pi T \vec{p}_i = \Pi T' \vec{p}_i$.*

Proof: Choose the 3-D affine transformation Q_i so that $Q_i P_i$ equals the $z = 0$ plane. (Recall that P_i is the plane containing V_i.) We can then consider the transformation $\Pi T Q_i^{-1}$ as a 2-D affine transformation when applied to $Q_i P_i$, that maps $Q_i V_i$ into R_i. Similarly, $\Pi T' Q_i^{-1}$ can be thought of as a 2-D affine transformation that maps $Q_i V_i$ inside R_i. Then, the region $Q V_i$ contains a fixed point under the transformation $(\Pi T Q_i^{-1})^{-1}(\Pi T' Q_i^{-1})$, which maps $Q V_i$ onto itself. So there exists some point, $\vec{q}_i \in Q_i V_i$ such that $\Pi T Q_i^{-1} \vec{q}_i = \Pi T' Q_i^{-1} \vec{q}_i$. Letting $\vec{p}_i = Q_i^{-1} \vec{q}_i$, then, we can see that $\Pi T \vec{p}_i = \Pi T' \vec{p}_i$, and that $\vec{p}_i \in V_i$. □

We may now use these lemmas to consider when T is uniquely determined. First, we point out that T is *not* uniquely determined in the case where a single line exists that intersects all regions. In this case, it is possible to view the model so that all regions intersect. As with the 2-D case, when all regions intersect, the forward constraints are satisfied by any affine transformation that shrinks the regions to a small area, that fits inside the intersection of the regions.

In particular, this tells us that when there are only two regions, a non-unique transformation is always possible. More generally, we may say that when the image regions are such that there is a non-identity, 2-D affine transformation mapping each region inside itself, that no set of model regions may be mapped uniquely to match these regions. That is, for regions such that $\Pi T V_i = R_i$, if there exists a 2-D affine transformation, $S \neq I$ (where I is the identity transformation) such that $S R_i \subseteq R_i$ for all i, then T cannot be unique. To see this, let:

$$Sp = \begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix} p + \begin{pmatrix} s_x \\ s_y \end{pmatrix}$$

and let:

$$\bar{S} p = \begin{pmatrix} s_{11} & s_{12} & 0 \\ s_{21} & s_{22} & 0 \\ 0 & 0 & 1 \end{pmatrix} p + \begin{pmatrix} s_x \\ s_y \\ 0 \end{pmatrix}.$$

Then:

$$\Pi \bar{S} T V_i = S \Pi T V_i \subseteq R_i$$

while at the same time, $\bar{S} T \neq T$. However, even though any pair of regions may produce an image that leads to non-unique solutions, it is still an open question whether they may also produce images that lead to unique solutions.

We consider next the general case where there are four regions that are not intersected by a plane. The case in which a plane exists that intersects all four or more regions is similar to the case in which there are three regions, and will be discussed later.

Theorem 7. *Assume the above definitions, with four regions, such that no single plane intersects all four regions. Then,* ΠT *is unique.*

Proof: By Lemma 3 there exists a point $\vec{p}_i \in V_i$ such that $\Pi T'(\vec{p}_i) = \Pi T(\vec{p}_i)$ for every $i = 0, \ldots, 3$. The points $\vec{p}_0, \ldots, \vec{p}_3$ are all non-coplanar. Consequently, since correspondences of four non-coplanar points determine a 3-D to 2-D affine transformation uniquely then $\Pi T = \Pi T'$. □

We now consider the case where there are only three regions, or there are four or more regions intersected by a plane, but the regions may not be intersected by any line. As before, let $\vec{p}_i \in V_i$ be points such that $\Pi T \vec{p}_i = \Pi T' \vec{p}_i$. Suppose this transformation is not unique, i.e., $T' V_i \subseteq R_i$ and T' not equivalent to T.

Using Lemma 2, we may assume WLOG that the model has been transformed so that $T p_i = p_i$, for $1 \leq i \leq 3$, and so that $T(0, 0, 1) = (0, 0, 1)$. This implies that $T = I$. This also implies that $T' \vec{p}_i = \vec{p}_i$, for $1 \leq i \leq 3$, so we may also assume, WLOG, that $\vec{p}_i$ is fixed under T', and hence that the $z = 0$ plane is fixed under T'. This tells us that we may write:

$$T' = \begin{pmatrix} 1 & 0 & k_1 \\ 0 & 1 & k_2 \\ 0 & 0 & k_3 \end{pmatrix}$$

Now define l_i to be the line of intersection of the planes P_i and $z = 0$. Let $R_i \cap l_i$ be the points $\vec{p}_i^{\,1}, \vec{p}_i^{\,2}$ (we will later consider the case where l_i intersects R_i in a single point). Consider one of these intersection points, $\vec{p}_i^{\,j}$. Let the tangent to R_i at $\vec{p}_i^{\,j}$ be $\vec{w} = (w_x, w_y, 0)$. Let the tangent to V_i at the point $\vec{p}_i^{\,j}$ have the direction $\vec{v}$. Then the directions of $\vec{w}$, $\Pi T \vec{v}$ and $\Pi T' \vec{v}$ must all be the same. Since T is the identity transformation, we must have $\vec{v} = (w_x, w_y, v_z)$ for some v_z. The points $\vec{p}_i^{\,j}$ are also fixed under T', since they lie in the $z = 0$ plane which is fixed under T'. Therefore, the tangent to $\Pi T' V_i$ at $\vec{p}_i^{\,j}$ is $(w_x + k_1 v_z, w_y + k_2 v_z)$. The condition that $T' V_i \subseteq R_i$ implies that $\Pi T' \vec{v}$ must have the same direction as $\vec{w}$. This implies that the directions of (w_x, w_y) and (k_1, k_2) must be parallel. The alternative, that $k_1 = k_2 = 0$ would imply that T is equivalent to T'. If the transformations are not equivalent, therefore, the tangents to each region R_i at a point $\vec{p}_i^{\,j}$ must all be parallel to (k_1, k_2), and so they must all be parallel to each other.

We now consider the possibility that the $z = 0$ plane intersects some regions, V_i in only a single point. If this is true, the remainder of V_i is either entirely above or below the $z = 0$ plane. Assume WLOG that it is above it. Then $\Pi T'$ maps any point, (x, y, z) that is on the boundary of V_i, to $(x, y) - z(k1, k2)$. Since T is the identity transform, (x, y) is on the boundary of R_i. Hence, T' displaces all points on the boundary of V_i in the same direction relative to the position to which ΠT maps them. Clearly some of these points will be mapped outside of R_i. Therefore, T not equivalent to T' implies that the $z = 0$ plane must intersect each region in two points, with opposite tangent directions.

We may now list some necessary conditions for T to be non-unique.

1. There exists a plane P, which intersects each region in two points.
2. For all points in the intersection of P and a region, the direction of tangency to the projections of the region 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, as in the case discussed in Section 3.4 for planar regions that are polygonal. In this case, the possible ranges of the tangent vectors must intersect.

This result shows that in general, three planar model regions will lead to a unique solution. To see this, note that we have three degrees of freedom in selecting a plane that intersects the model regions. The conditions above indicate that a degeneracy will occur when six tangent vectors are parallel after we project them onto this plane, in the direction of the true viewing direction. This leads to five constraints. Since we have more independent constraints than degrees of freedom with which to satisfy them, in general the conditions will not be satisfied.

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 (Canny, 1983). We then constructed polygonal approximations to the edges using Pavlidis and Horowitz's (1974) 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 (Jacobs, 1996). The matching between the regions was specified manually. Finally, the transformations relating these images were recovered using either the forward or backward solutions. The

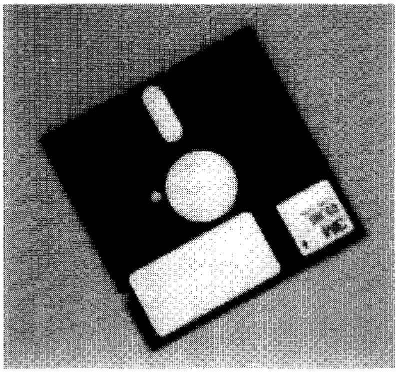

Figure 10. An image of a computer diskette used as a model.

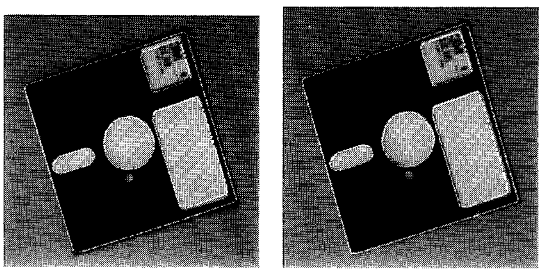

Figure 11. Matching the diskette model to a novel image of the diskette under similarity (left figure, $\lambda = -1.89$) and affine (right, $\lambda = -1.27$) transformations.

figures below contain overlay images of the aligned model and image. We also provide the value of λ obtained in every experiment. In all of the experiments we obtained negative values for λ. This is because sensing errors often cause slight violations of the one-way constraints. A near perfect alignment between the models and the images is achieved nevertheless in accordance to the theory developed in the paper.

Figure 10 shows an image of a diskette used as a model. Figure 11 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 12 shows the result of matching the model to the same image using only two regions. Figure 13 shows 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.

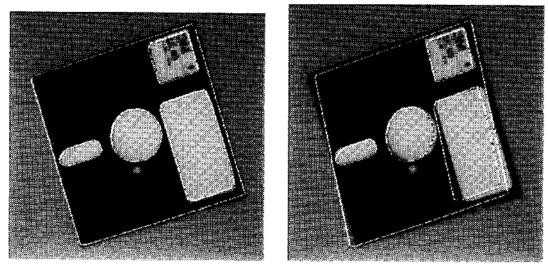

Figure 12. Matching the diskette model to a novel image of the diskette using two regions only (the left and the upper right, left figure: similarity, $\lambda = -1.56$, right: affine, $\lambda = -0.69$).

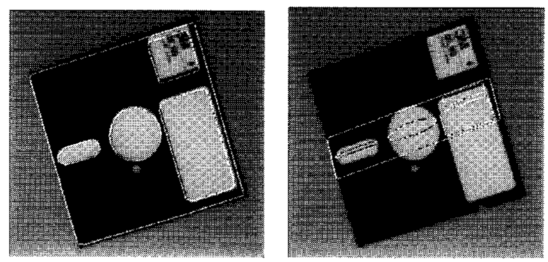

Figure 13. Matching the diskette model to a novel image of the diskette using two degenerate regions only (the left and the lower right, left figure: similarity, $\lambda = -1.15$, right: affine, $\lambda = -0.55$).

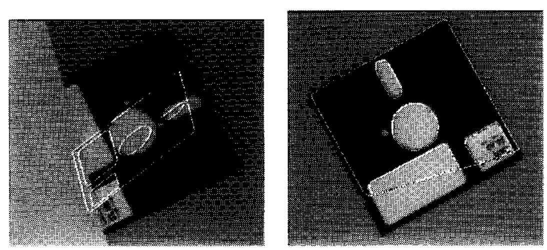

Figure 14. Matching the diskette model to a novel image that contains unknown occlusion under an affine transformation using the forward (left figure, $\lambda = -12.59$) and the backward (right, $\lambda = -1.51$) constraints.

Figures 14 and 15 demonstrate the performance of the system in the presence of unknown partial occlusion. When the forward constraints were used (left images) an over contraction of the model was obtained since these constraints are inconsistent with the presence of unknown occlusion. To obtain better results one has to first eliminate the constraints which are due to the occlusion (by identifying the locations of occlusion). When the backward constraints were used (right images) in both the similarity and affine cases a good match was obtained. In this image, three of the five regions are occluded. Since in the affine case the

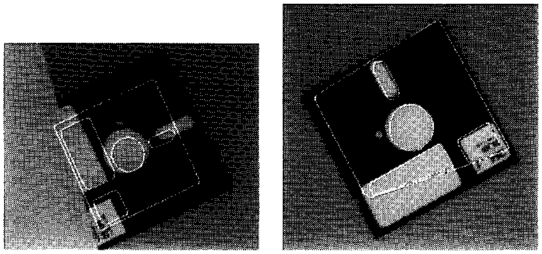

Figure 15. Matching the diskette model to a novel image that contains unknown occlusion under a similarity transformation using the forward (left figure, $\lambda = -15.60$) and the the backward (right, $\lambda = -3.08$) constraints.

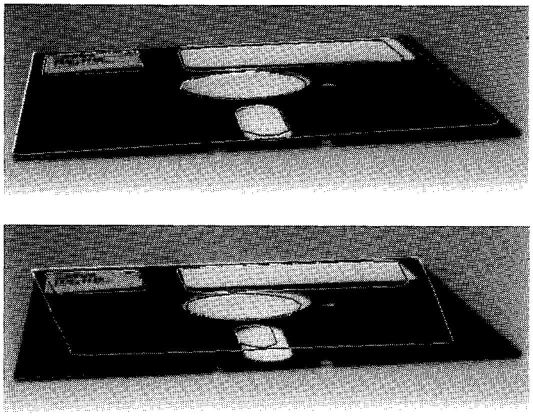

Figure 16. Matching the diskette model to a novel image containing relatively large perspective distortions under projective (top figure, $\lambda = -2.25$) and affine (bottom, $\lambda = -5.38$) transformations.

remaining two regions are degenerate by themselves, the partial information obtained from the occluded regions is essential to producing an accurate result.

Figure 16 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 17 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, Fig. 18 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 results are shown in Fig. 19.

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.

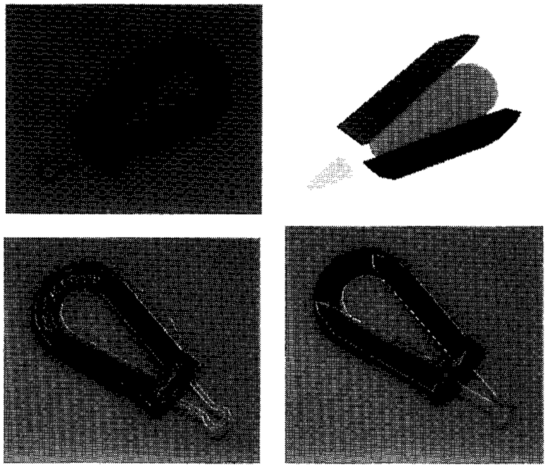

Figure 17. Matching a model of a magnet to a novel image: the model (top left), the regions extracted from the model (top right), the match (under affine transformation, bottom left, $\lambda = -3.46$), and the overlayed regions (bottom right).

Figure 18. Two images of a book.

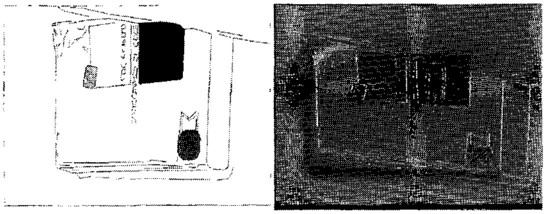

Figure 19. Matching the two images of the book under a 2-D affine transformation. The three regions extracted (left image, the regions are shaded) and the match obtained (right, $\lambda = -3.82$).

6. Conclusion

We have presented a fundamentally new approach to the pose determination part of the object recognition problem. Perhaps what is most novel about our approach is the weaker requirements that it makes on correspondence, compared to previous approaches. Local methods explicitly require a correspondence between simple local features such as points and lines before determining pose. Global methods implicitly produce such a correspondence as well. Moment based methods, for example, compute points (such as center of mass) or lines (such as moments of inertia) from regions, and determine pose based on such correspondences. Our method, while still requiring a correspondence between regions, does not require an explicit correspondence between local features before determining pose.

It is well-known that past methods have some drawbacks associated with their need for correspondences between local image and model properties. The detection of local features, such as corners and lines, can be highly sensitive to noise and viewpoint variation because these features do not reflect the overall shape of an object, but instead capture properties of a small portion of an object's boundary. Global features of a region, such as its center of mass, can be much more resistant to noise, but may be highly sensitive to occlusion. (In fact, depending on a region's shape, its higher order moments may also be sensitive to noise.) When we have hypothesized a correspondence between two regions, we would prefer not to have to further hypothesize a correspondence between their moments, or to find and match local features of their boundaries. Rather, if possible we would like to make use of a more minimal assumption; that the image region was produced by the model region. Our one-way constraints make use of only this minimal assumption.

Naturally, if we can infer more detail in a correspondence, and match specific points or lines of a model and region, it is useful to take advantage of this information, and our approach allows us to take full advantage of this knowledge when it is present. But it also shows how to find pose from a much weaker statement. 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. For example, it is fundamental to point-based pose determination to know that a correspondence between three points determines a finite number of poses, under scaled-orthographic projection. Similarly, it is fundamental to our approach to know that a correspondence between two coplanar regions or three non-coplanar regions generally determines a unique pose, under scaled-orthographic projection. 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.

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 shape and local boundary structure when available.

Acknowledgments

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

References

Alter, T.D. and Grimson, W.E.L. 1993. Fast and robust 3D recognition by alignment. In *Proc. Fourth Inter. Conf. Computer Vision*, pp. 113–120.

Alter, T.D. and Jacobs, D. 1994. Error propagation in full 3D-from-2D object recognition. *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 892–898.

Amenta, N. 1994. Bounded boxes, Hausdorff distance, and a new proof of an interesting Helly-type theorem. *Proceedings of the 10th Annual ACM Symposium on Computational Geometry*, pp. 340–347.

Ayache, N. and Faugeras, O. 1986. HYPER: A new approach for the recognition and positioning of two-dimensional objects. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 8(1):44–54.

Baird, H. 1985. *Model-Based Image Matching Using Location*. MIT Press: Cambridge.

Bajcsy, R. and Solina, F. 1987. Three dimensional object representation revisited. *Proc. of The First Int. Conf. on Computer Vision*, London, pp. 231–240.

Basri, R. 1996. Paraperspective ≡ Affine. *Int. J. of Comp. Vis.*, 19(2):169–179.

Basri, R. and Ullman, S. 1993. The alignment of objects with smooth surfaces. *Computer Vision, Graphics, and Image Processing: Image Understanding*, 57(3):331–345.

Basri, R. and Jacobs, D.W. 1994. Recognition using region correspondences. *The Weizmann Institute of Science*, T.R. CS95-33.

Basri, R. and Jacobs, D.W. 1995. Recognition using region correspondences. *IEEE Int. Conf. on Computer Vision*, pp. 8–15.

Biederman, I. 1985. Human image understanding: Recent research and a theory. *Computer Graphics, Vision, and Image Processing*, 32:29–73.

Binford, T. 1971. Visual perception by computer. *IEEE Conf. on Systems and Control*.

Breuel, T. 1991. Model based recognition using pruned correspondence search. *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 257–268.

Brooks, R. 1981. Symbolic reasoning among 3-D models and 2-D images. *Artificial Intelligence*, 17:285–348.

Canny, J. 1983. A computational approach to edge detection. *Trans. on Pattern Recognition and Machine Intelligence*, 8(6):679–698.

Cass, T. 1992. Polynomial time object recognition in the presence of clutter, occlusion and uncertainty. *Second European Conf. on Computer Vision*, pp. 834–842.

Conway, J.B. 1990. *A Course in Functional Analysis*. Springer-Verlag.

Coxeter, H.S.M. 1993. *The Real Projective Plane*. Springer-Verlag.

Darrell, T., Sclaroff, S., and Pentland, A. 1990. Segmentation by minimal description. *IEEE Int. Conf. on Computer Vision*, Japan, pp. 112–116.

Duda, R.O. and Hart, P.E. 1973. *Pattern Classification and Scene Analysis*. Wiley-Interscience Publication, John Wiley and Sons, Inc.

Dudani, S.A., Breeding, K.J., and McGhee, R.B. 1977. Aircraft identification by moments invariants. *IEEE Trans. on Computations*, C-26(1):39–46.

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 A.C.M.*, 24(6):381–395.

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 Pattern Analysis and Machine Intelligence*, 13(10):971–991.

Hoffman, D.D. and Richards, W. 1985. Parts of recognition. *Cognition*, 18:65–96.

Horaud, R. 1987. New methods for matching 3-D objects with single perspective views. *IEEE Trans. Pattern Anal. Machine Intell.*, 9(3):401–412.

Hu, M.K. 1962. Visual pattern recognition by moment invariants. *IRE Trans. on Information Theory*, IT-8:169–187.

Huttenlocher, D.P. and Ullman, S. 1990. Recognizing solid objects by alignment with an image. *Int. J. Computer Vision*, 5(2):195–212.

Huttenlocher, D., Klanderman, G., and Rucklidge, W. 1993a. Comparing images using the Hausdorff distance. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 15(9):850–863.

Huttenlocher, D., Noh, J., and Rucklidge, W. 1993b. Tracking nonrigid objects in complex scenes. *4th Int. Conf. on Computer Vision*, pp. 93–101.

Jacobs, D. 1992. Space efficient 3D model indexing. *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 439–444.

Jacobs, D. 1996. Robust and efficient detection of convex groups. *IEEE Trans. PAMI* (18)1:23–37.

Joshi, T., Ponce, J., Vijayakumar, B., and Kriegman, D. 1994. Hot curves for modelling and recognition of smooth curved 3D objects. *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 876–880.

Koenderink, J. and van Doorn, A. 1991. Affine structure from motion. *Journal of the Optical Society of America*, 8(2):377–385.

Kriegman, D. and Ponce, J. 1990. On recognizing and positioning curved 3-D objects from image contours. *IEEE Trans. Pattern Anal. Machine Intell.*, 12(12):1127–1137.

Lamdan, Y. and Wolfson, H.J. 1988. Geometric hashing: A general and efficient model-based recognition scheme. *Second International Conf. Computer Vision*, pp. 238–249.

Lamdan, Y., Schwartz, J.T., and Wolfson, H.J. 1990. Affine invariant model-based object recognition. *IEEE Trans. Robotics and Automation*, 6:578–589.

Lowe, D. 1985. *Perceptual Organization and Visual Recognition*. Kluwer Academic Publishers: The Netherlands.

Marr, D. and Nishihara, H. 1978. Representation and recognition of the spatial organization of three dimensional structure. *Proceedings of the Royal Society of London B*, 200:269–294.

Pavlidis, T. and Horowitz, S. 1974. Segmentation of plane curves. *IEEE Trans. on Computers*, TC-23:860–870.

Pentland, A. 1987. Recognition by parts. *Proceedings of the First International Conf. on Computer Vision*, pp. 612–620.

Persoon, E. and Fu, K.S. 1977. Shape descimination using Fourier descriptors. *IEEE Trans. on Systems, Man and Cybernetics*, 7:534–541.

Poelman, C.J. and Kanade, T. 1994. A paraperspective factorization method for shape and motion recovery. *Proc. of European Conf. on Computer Vision*.

Reeves, A.P., Prokop, R.J., Andrews, S.E., and Kuhl, F.P. 1988. Three-dimensional shape analysis using moments and Fourier descriptors. *Trans. on Pattern Recognition and Machine Intelligence*, 10(6):937–943.

Richard, C.W. and Hemami, H. 1974. Identification of three dimensional objects using Fourier descriptors of the boundry curve. *IEEE Trans. on Systems, Man and Cybernetics*, 4(4):371–378.

Rothwell, C., Zisserman, A., Mundy, J., and Forsyth, D. 1992. Efficient model library access by projectively invariant indexing functions. *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 109–114.

Rothwell, C.A., Zisserman, A., Forsyth, D.A., and Mundy, J.L. 1992. Canonical frames for planar object recognition. *Proc. of 2nd Eur. Conf. on Computer Vision*, pp. 757–772.

Sadjadi, F.A. and Hall, E.L. 1980. Three-dimensional moment invariants. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2(2):127–136.

Subrahmonia, J., Cooper, D., and Keren, D. 1996. Practical reliable Bayesian recognition of 2D and 3D objects using implicit polynomials and algebraic invariants. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(5):505–519.

Sugimoto, A. 1996. Object recognition by combining paraperspective images. *Int. J. of Comp. Vis.*, 19(2):181–201.

Thompson, D.W. and Mundy, J.L. 1987. Three dimensional model matching from an unconstrained viewpoint. *Proc. of IEEE Int. Conf. on Robotics and Automation*, pp. 208–220.

Tomasi, C. and Kanade, T. 1992. Shape and motion from image streams under orthography: A factorization method. *International Journal of Computer Vision*, 9(2):137–154.

Ullman, S. and Basri, R. 1991. Recognition by linear combinations of models. *IEEE Trans. on PAMI*, 13(10):992–1006.

Weiss, I. 1993. Geometric invariants and object recognition. *International Journal of Computer Vision*, 10(3):207–231.