Robust Affine Structure Matching for 3D Object Recognition

Todd A. Cass

Xerox Palo Alto Research Center
Palo Alto, CA 94304 USA, cass@parc.xerox.com

Abstract. This paper considers a model-based approach to identifying and locating known 3D objects from 2D images. The method is based on geometric feature matching of the model and image data, where both are represented in terms of *local geometric features*. This paper extends and refines previous work on feature matching using transformation constraint methods by detailing the case of full 3D objects represented as point features and developing geometric algorithms based on *conservative approximations* to the previously presented general algorithm which are much more computationally feasible.
Keywords: *Object Recognition, Geometric Feature Matching, Pose Recovery, Pose Refinement, Affine Structure, Model Indexing, Computational Geometry.*

1 Introduction

This paper considers the task of identifying and locating known 3D objects in 2D images of a scene containing them by matching geometric representations of a model of the object to geometric representations of the image data—a task often called *model-based object recognition*. We take a *model-based* approach to *interpreting* subsets of sensory data corresponding to *instances* of objects in the scene in terms of *models* of the objects known *a priori*. Both the model and the image data are represented in terms of sets of **local geometric features**[1] consisting of geometric elements like points, line segments, and curve normals. An approach based on **geometric feature matching** is used to interpret the image by finding both the model to data feature correspondences and a 3D pose of the model which could produce the observed instance.

Geometric feature matching is difficult because the correspondences between the model features and the corresponding subset of data features are unknown, there is clutter from data features with no model correspondence, there are model features which have no corresponding data feature due to occlusion and other processes, and there is geometric uncertainty in the measurement of the data features arising from the imaging process, the feature representation process, and scene events such as occlusion.

[1] Throughout this paper we will follow the convention that terms being effectively defined will appear in sans-serif **bold**, while items being emphasized for clarity will appear in *italics*.

This paper considers the case of 3D objects and 3D transformations, greatly extending and refining previously reported techniques for geometric matching based on discrete constraints and methods from computational geometry first presented for the case of 2D objects and planar transformations and 2D and 3D objects and 3D transformations[5, 6, 8].

We present a technique for feature matching based on analyzing constraints on feasible feature matchings in transformation space. It applies to affine feature matching of 2D and 3D point sets and 2D and 3D affine motion. The main contribution is a method for carefully projecting the constraints in the high-dimensional transformation space, an 8-dimensional space for 3D affine motion, down to a two dimensional transformation space. This approximation is conservative in that it will never exclude feasible matches. This approximate projection then allows the use of planar geometric algorithms (which is very important due to the great difficulty of geometric computations in higher dimensions) for analyzing these transformation constraints to find feasible matchings. Efficient algorithms for this planar geometric analysis are presented, and experimental results of their application to matching are presented.

1.1 The Case of 3D Models and 2D Data

This section will first give a brief outline of some of the notation used and the key ideas of this geometric feature matching method; a more detailed explanation can be found in other papers[6, 7, 9]. The local geometric features used here consist of points. Denote a model feature $\mathbf{m}$ by a vector representing the feature's position and denote the position of a data feature by $\mathbf{d}$. Denote the bounded uncertainty[3, 13] region U^i for the data feature position measurement, for feature $\tilde{\mathbf{d}}_i$. We let $\tilde{\mathbf{d}}_i$ denote the measured value of a model feature and $\mathbf{d}_i$ denote the value without error. The true position of $\mathbf{d}_i$ falls in the set U^i, which contains $\tilde{\mathbf{d}}_i$.

A single transformed model feature is aligned with a data feature when $T[\mathbf{m}_i] \in U^j$, so the set of transformations T geometrically feasible for a particular model and data feature match $(\mathbf{m}_i, \tilde{\mathbf{d}}_j)$ is the set of transformations $\mathcal{F}^{ij} = \{T \in \Omega | T[\mathbf{m}_i] \in U^j\}$. Define a match as a pairing of a model feature and a data feature, and a match set, denoted $\mathcal{M}$, as an arbitrary set of matches—some subset of all possible matches. A match set is called geometrically consistent if and only if $\bigcap_{(i,j)} \mathcal{F}^{ij} \neq \emptyset$ for (i, j) such that $(\mathbf{m}_i, \mathbf{d}_j) \in \mathcal{M}$.

With each transformation T we associate the function $\varphi(T) = \{(\mathbf{m}_i, \tilde{\mathbf{d}}_j) | T[\mathbf{m}_i] \in U^j\}$ defined to be the set of feature matches aligned by the transformation T. We call the match set $\varphi(T)$ a maximal geometrically-consistent match set. A match set $\mathcal{M}$ is a maximal geometrically-consistent match set (or, a maximal match set) if it is the largest geometrically consistent match set at some transformation T. Maximal match sets are the only sets of feature correspondences we need to consider, and in the cases of interest here there are only a polynomial number of them.

Finally, we define the notion of a geometric consistency measure, $\Phi(T)$, characterizing the number of matches feasible for a given transformation T. The simplest measure is a count of the number of matches feasible at a transformation $\Phi^{\mathcal{M}}(T) = |\varphi(T)|$; we use a more sophisticated measure which is the minimum of the number of distinct model features and distinct data features appearing in the feasible match set[17] given by

$$\Phi_T(T) = \mathtt{min}\left(\left|\{\mathbf{m}_i; (\mathbf{m}_i, \tilde{\mathbf{d}}_k) \in \varphi(T)\}\right|, \left|\{\mathbf{d}_j; (\mathbf{m}_l, \tilde{\mathbf{d}}_j) \in \varphi(T)\}\right|\right) \qquad (1)$$

which approximates the size of a one-to-one feature correspondence between the model and a data subset.

This paper will focus on using linear bounds on geometric uncertainty. This formulation has been applied to 2D models and transformations and 2D data[3, 4, 7] and 2D models and data and 3D transformations[6, 14]. Here we consider 3D models and transformations and 2D data. The 3D rigid transformation and scaled orthographic projection mapping a model point $\mathbf{m}'_i \in \mathbb{R}^3$ onto a data point $\mathbf{d}'_j \in \mathbb{R}^2$ can be expressed by the transformation $\mathbf{S}\mathbf{m}'_i + \mathbf{t} = \mathbf{d}'_j$ where $\mathbf{S} = \sigma\mathbf{P}\mathbf{R} = \begin{bmatrix} s_{11} & s_{12} & s_{13} \\ s_{21} & s_{22} & s_{23} \end{bmatrix}$ with $0 < \sigma \in \mathbb{R}$ a positive scale factor, $\mathbf{P}$ a 2×3 projection matrix simply eliminating the third row of $\mathbf{R}$, where $\mathbf{R} \in SO_3$, and $\mathbf{t}, \mathbf{d}'_j \in \mathbb{R}^2$. From $\mathbf{S} = \sigma\mathbf{P}\mathbf{R}$ and the fact the $\mathbf{R}$ is an orthonormal matrix we have the quadratic constraint that

$$\mathbf{S}\mathbf{S}^t = \sigma^2(\mathbf{P}\mathbf{R})(\mathbf{R}^t\mathbf{P}^t) = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}. \qquad (2)$$

For convenience we will refer to $\mathbf{S}$ as the linear component of the transformation, and $\mathbf{t}$ the translation component.

In essence the approach is to define a model for geometric uncertainty in the data features which in turn imposes constraints on the geometrically feasible transformations aligning model and data feature sets. This frames geometric feature matching as a problem in computational geometry, specifically constructing and exploring arrangements of constraints on geometrically feasible (i.e. structurally consistent) matchings of model features to data features.

For each data feature formulate a convex polygonal uncertainty region bounding the measurement uncertainty consisting of the set of points $\mathbf{x}'$ characterized by a set of linear inequalities[3] $(\mathbf{x}' - \tilde{\mathbf{d}}'_j)^t\hat{\mathbf{n}}_l \geq -\epsilon^l$. For feature match $(\mathbf{m}'_i, \tilde{\mathbf{d}}'_j)$ this imposes constraints on feasible transformations of the form:

$$(\mathbf{S}\mathbf{m}'_i + \mathbf{t} - \tilde{\mathbf{d}}'_j)^t\hat{\mathbf{n}}_l \geq -\epsilon^l \qquad (3)$$

for $l = 1, ..., k$ where $\hat{\mathbf{n}}_l$ is the unit normal vector and ϵ_l the scalar distance describing each linear constraint for $l = 1, ..., k$. This describes a polytope of feasible transformation parameters, $\mathcal{F}^{ij}$, mapping $\mathbf{m}'_i$ to within the uncertainty bounds of $\tilde{\mathbf{d}}'_j$. Each possible feature match $(\mathbf{m}'_i, \tilde{\mathbf{d}}'_j)$ yields such a constraint. Feature matching can be accomplished by searching an arrangement of the boundaries of the polytopes $\mathcal{F}^{ij}$ for all i and j to find sets of transformation parameters with a high geometric consistency measure, i.e. satisfying many of the match constraints. Conceptually the constraints forming the boundaries of the polytopes $\mathcal{F}^{ij}$ partition the parameter space into cells. Exploring the arrangement then involves finding the cells with high geometric consistency.

In the case of 3D affine transformation and orthographic projection (where we relax the quadratic constraints) there are 8 transformation parameters. If there are M model features and N data features then the arrangement of the bounding constraints for each feasible polytopes $\mathcal{F}^{ij}$ has size $O(M^8N^8)$[11]; constructing and exploring it seems a formidable task. Nonetheless, an important point is that finding *all* geometrically feasible matchings of model and data features within the uncertainty bounds is a polynomial-time problem, in contrast to worst-case exponential-time approaches such as popular feature correspondence-space searches.

2 Approximating the Constraint Arrangement

2.1 The Image of 3D Points Under Scaled-Orthography

Unfortunately constructing and exploring arrangements of hyperplanes in high dimensional spaces is extremely difficult, and computationally infeasible in this case. There is, however, considerable structure to the feature matching problem which we can exploit to make the computations more feasible and to allow the introduction of careful approximations to greatly reduce the computation required to find large geometrically consistent match sets. One approach to this is to work directly with the constraints in the s-parameter space spanned by the coordinates s_{ij}. This approach was outlined for the 3D case previously[6]. This section introduces an elegant formulation which through convenient choice of representations and careful approximation reduces this from a constraint problem in an 8D parameter space to a constraint problem in a 2D parameters space which is much more computationally feasible.

Let $\{\mathbf{m}'_i\}_{i=0}^{M-1}$ be a set of 3D model points, and $\{\mathbf{d}'_i\}_{i=0}^{M-1}$ be the set of points forming their image without error from an arbitrary viewpoint under scaled-orthography such that $\mathbf{S}\mathbf{m}'_i + \mathbf{t} = \mathbf{d}'_i$. Distinguish three model points $\mathbf{m}'_{\pi_0}$, $\mathbf{m}'_{\pi_1}$, and $\mathbf{m}'_{\pi_2}$ and their corresponding image points $\mathbf{d}'_{\pi_0}$, $\mathbf{d}'_{\pi_1}$, and $\mathbf{d}'_{\pi_2}$ which we will call primary model features and primary data features, respectively. Define $\mathbf{m}'_i \neq \mathbf{m}'_{\pi_0}, \mathbf{m}'_{\pi_1}, \mathbf{m}'_{\pi_2}$ and $\mathbf{d}'_i \neq \mathbf{d}'_{\pi_0}, \mathbf{d}'_{\pi_1}, \mathbf{d}'_{\pi_2}$ secondary model features and secondary data features, respectively. Finally, define $\mathbf{m}_i = (\mathbf{m}'_i - \mathbf{m}'_{\pi_0})$ and $\mathbf{d}_i = (\mathbf{d}'_i - \mathbf{d}'_{\pi_0})$ for all i which defines new coordinate frames for the model and the image data with $\mathbf{m}'_{\pi_0}$ and $\mathbf{d}'_{\pi_0}$ as origins, respectively. The new primary features are thus $\mathbf{m}_{\pi_1}$, $\mathbf{m}_{\pi_2}$, $\mathbf{d}_{\pi_1}$, and $\mathbf{d}_{\pi_2}$ and we now have $\mathbf{S}\mathbf{m}_i = \mathbf{d}_i$ for all i.

Next, express the model point $\mathbf{m}_i$ using $\mathbf{m}_{\pi_1}$, $\mathbf{m}_{\pi_2}$, and $\hat{\mathbf{n}} = \frac{\mathbf{m}_{\pi_1} \times \mathbf{m}_{\pi_2}}{|\mathbf{m}_{\pi_1} \times \mathbf{m}_{\pi_2}|}$ as basis vectors[2]. Assuming $\mathbf{m}_{\pi_1}$ and $\mathbf{m}_{\pi_2}$ are linearly independent there exist three constants a_i, b_i, and c_i such that
$\mathbf{m}_i = a_i\mathbf{m}_{\pi_1} + b_i\mathbf{m}_{\pi_2} + c_i\hat{\mathbf{n}} = [\,\mathbf{m}_{\pi_1} \quad \mathbf{m}_{\pi_2}\,] \begin{bmatrix} a_i \\ b_i \end{bmatrix} + c_i\hat{\mathbf{n}}$. When the three primary model and data features are exactly aligned by some linear transformation we

have $\mathbf{S}\,[\,\mathbf{m}_{\pi_1} \quad \mathbf{m}_{\pi_2}\,] = [\,\mathbf{d}_{\pi_1} \quad \mathbf{d}_{\pi_2}\,]$. In this case it can be shown[9] that the image of any secondary model feature $\mathbf{m}_i$ under the imaging transformation $\mathbf{S}$ which aligns the primary features (see also Jacobs[18]) can be expressed:

$$\mathbf{S}\mathbf{m}_i = c_i \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + [\,\mathbf{d}_{\pi_1} \quad \mathbf{d}_{\pi_2}\,] \begin{bmatrix} a_i \\ b_i \end{bmatrix} \tag{4}$$

where ξ_1 and ξ_2 are parameters completely determining the viewpoint of the model. The image of $\mathbf{m}_i$ under the matrix $\mathbf{S}$ thus has two components. The component due to the planar component of $\mathbf{m}_i$, given by $[\,\mathbf{d}_{\pi_1} \quad \mathbf{d}_{\pi_2}\,]\,[\,a_i \quad b_i\,]^t$, and the component due to the non-planar component of $\mathbf{m}_i$, given by $c_i\,[\,\xi_1 \quad \xi_2\,]^t$.

Let's call equation (4) the reprojection equation for a secondary model feature $\mathbf{m}_i$ with respect to the primary feature alignments. Note that by choosing $\mathbf{m}'_{\pi_0}$ and $\mathbf{d}'_{\pi_0}$ as origins we have eliminated the translation component of the transformation, for now.

When the transformation given by $\mathbf{S}$ and $\mathbf{t}$ corresponds to 3D affine transformation and orthographic projection, the vector $[\,\xi_1 \quad \xi_2\,]^t$ can take on any value corresponding to different viewpoints and affine transformations. When the correspondence between the primary features is used to compute a *rigid* alignment transformation then the components $[\,\xi_1 \quad \xi_2\,]^t$ must satisfy two quadratic constraints, having at most two solutions[16, 2] yielding rigid 3D motion and scaled orthographic projection. This follows directly[10] from equations (2) and (4) and is closely related to similar quadratic constraints in other formulations[16, 2].

2.2 Incorporating the Effects Of Uncertainty

Central to the reprojection equation and the computation of the rigid three-point alignment transformation are the three primary model and data feature matches. The effect of geometric uncertainty in these primary features is significant and must be handled carefully, in addition to uncertainty in the secondary model features. We will follow an approach representing the measured data feature and the error as explicit terms[12]. When there is geometric uncertainty in the measured position of the data features the transformation mapping model feature $\mathbf{m}'_i$ to the measured data feature $\tilde{\mathbf{d}}'_j$ must satisfy $\mathbf{S}\mathbf{m}'_i + \mathbf{t} = \tilde{\mathbf{d}}'_j = \mathbf{d}'_j + \mathbf{e}'_j$, where again $\mathbf{d}'_j$ represents the (unknown and unmeasurable) true position of the data feature, $\mathbf{e}'_j$ the measurement error, and $\mathbf{d}'_j + \mathbf{e}'_j = \tilde{\mathbf{d}}'_j$ represents the *observed* location of the data feature. As before we match $\mathbf{m}'_{\pi_0}$ and $\mathbf{d}'_{\pi_0}$ as corresponding origins in the model and data coordinate frames, respectively, and define $\mathbf{m}_i = (\mathbf{m}'_i - \mathbf{m}'_{\pi_0})$ for all i and $\tilde{\mathbf{d}}_j = (\tilde{\mathbf{d}}'_j - \tilde{\mathbf{d}}'_{\pi_0})$ for all j. We then have $\mathbf{S}\mathbf{m}_i = \tilde{\mathbf{d}}_j = \mathbf{d}_j + \mathbf{e}_j$ where we define $\mathbf{e}_j = \mathbf{e}'_j - \mathbf{e}'_{\pi_0}$.

When the reprojection or alignment transformation is computed from the *measured* geometry of the primary data features $\tilde{\mathbf{d}}_{\pi_1}$ and $\tilde{\mathbf{d}}_{\pi_2}$ the reprojection of a secondary model feature is given by $\mathbf{S}\mathbf{m}_i = c_i \begin{bmatrix} \tilde{\xi}_1 \\ \tilde{\xi}_2 \end{bmatrix} + [\,\tilde{\mathbf{d}}_{\pi_1} \quad \tilde{\mathbf{d}}_{\pi_2}\,] \begin{bmatrix} a_i \\ b_i \end{bmatrix}$ where in the case of rigid alignment $\tilde{\xi}_1$ and $\tilde{\xi}_2$ are computed from the noisy primary data features $\tilde{\mathbf{d}}_{\pi_1}$ and $\tilde{\mathbf{d}}_{\pi_2}$. In the affine case $\tilde{\xi}_1 = \xi_1$ and $\tilde{\xi}_2 = \xi_2$

can be anything. The planar component of the reprojection has a simple linear form[12] given by $[\,\tilde{\mathbf{d}}_{\pi_1} \quad \tilde{\mathbf{d}}_{\pi_2}\,] \begin{bmatrix} a_i \\ b_i \end{bmatrix} = [\,\mathbf{d}_{\pi_1} \quad \mathbf{d}_{\pi_2}\,] \begin{bmatrix} a_i \\ b_i \end{bmatrix} + [\,\mathbf{e}_{\pi_1} \quad \mathbf{e}_{\pi_2}\,] \begin{bmatrix} a_i \\ b_i \end{bmatrix}$. The non-planar component of the reprojection has a complicated non-linear dependence on the uncertainty in the primary data features[10] which is a problem when computing $\tilde{\xi}_1$ and $\tilde{\xi}_2$ for a rigid transformation. We will only be concerned with 3D affine transformation, and will constrain ξ_1 and ξ_2 without explicitly computing them.

3 Geometric Feature Matching

We now consider a sequence of formulations which make successively more approximations to reduce the dimensionality of the geometric constraints. We emphasize that these approximations are *conservative* in that they will never exclude a feasible transformation. They are also careful approximations in that they do not include *too many* infeasible transformations. These formulations illustrate that the geometric feature matching problem has both a purely combinatorial component as well as a geometric component and the different balance between these components.

An 8D Constraint Problem: Starting with the equation for an aligned feature match incorporating error $\mathbf{S}\mathbf{m}'_i + \mathbf{t} - \tilde{\mathbf{d}}'_j = -\mathbf{e}'_j$, from the definition of the uncertainty bounds we have $\hat{\mathbf{n}}_l^t \mathbf{e}'_p \leq \epsilon^l$, for $l = 1, ..., k$ and for all p, which directly leads to the inequalities (3). These inequalities describe an 8D constraint arrangement of size $O(M^8 N^8)$ which can be explored in $O(M^8 N^8)$ time[11].

A 6D Constraint Problem: Next hypothesize the correspondence between one model feature $\mathbf{m}'_{\pi_0}$ and one data feature $\tilde{\mathbf{d}}'_{\rho_0}$ and let these points be the origin for the model and image coordinate frames, respectively. We subscript the primary data feature $\tilde{\mathbf{d}}_{\rho_0}$ differently than the model feature $\mathbf{m}_{\pi_0}$ to emphasize that their correspondence to one another is hypothesized, not known. As before we define $\mathbf{m}_i = (\mathbf{m}'_i - \mathbf{m}'_{\pi_0})$ and $\tilde{\mathbf{d}}_i = (\tilde{\mathbf{d}}'_i - \tilde{\mathbf{d}}'_{\rho_0})$ for all i, and the transformation aligning $\mathbf{m}_i$ and $\tilde{\mathbf{d}}_j$ must satisfy $\mathbf{S}\mathbf{m}_i - \mathbf{d}_j = 0$ or equivalently $\mathbf{S}\mathbf{m}_i - \tilde{\mathbf{d}}_j = -\mathbf{e}_j$ where we define $\mathbf{e}_j = \mathbf{e}'_j - \mathbf{e}'_{\rho_0}$. Using again the uncertainty bound $\hat{\mathbf{n}}_l^t \mathbf{e}'_p \leq \epsilon^l$, for $l = 1, ..., k$ and for all p we then have constraints for a single match on the 6 s-parameters which comprise the linear component of the transformation[6]: $(\mathbf{S}\mathbf{m}_i - \tilde{\mathbf{d}}_j)^t \hat{\mathbf{n}}_l \geq -2\epsilon^l$ for $l = 1, ..., k$. This constraint arrangement is of size $O(M^6 N^6)$, and we see that we have shifted around where the computation must be done. For each choice of π_0 and ρ_0 we analyze the arrangement of s-parameter constraints, of size $O(M^6 N^6)$, but must perform a combinatorial search through the MN possible different primary matches $(\mathbf{m}_{\pi_0}, \tilde{\mathbf{d}}_{\rho_0})$ for $\pi_0 = 1, ..., M-1$, $\rho_0 = 1, ..., N-1$, and for each match do the s-parameter constraint analysis described above. This is also an approximation because each uncertainty region shares the error vector $\mathbf{e}_{\rho_0}$ which can only take on one global value. Thus we are computing approximate geometric consistency, but conservatively in the sense that the bounds are conservative thus we will never underestimate true global consistency in the arrangement.

A 2D Constraint Problem: Next assume the correspondence between three primary model and data features, $(\mathbf{m}'_{\pi_0}, \tilde{\mathbf{d}}_{\rho_0})$, $(\mathbf{m}'_{\pi_1}, \tilde{\mathbf{d}}_{\rho_1})$, and $(\mathbf{m}'_{\pi_2}, \tilde{\mathbf{d}}_{\rho_2})$, and represent the model features in terms of the natural affine coordinate frame as discussed in section 2.1. In order for a feature match $(\mathbf{m}_i, \mathbf{d}_j)$ to be feasible we must have that $c_i \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + [\,\mathbf{d}_{\pi_1} \quad \mathbf{d}_{\pi_2}\,] \begin{bmatrix} a_i \\ b_i \end{bmatrix} = \mathbf{d}_j$ for some value of the viewpoint parameters ξ_1 and ξ_2. Using the actual observed data features and accounting for uncertainty we must have $c_i \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + [\,\tilde{\mathbf{d}}_{\pi_1} \quad \tilde{\mathbf{d}}_{\pi_2}\,] \begin{bmatrix} a_i \\ b_i \end{bmatrix} - \tilde{\mathbf{d}}_j = [\,\mathbf{e}_{\pi_1} \quad \mathbf{e}_{\pi_2}\,] \begin{bmatrix} a_i \\ b_i \end{bmatrix} - \mathbf{e}_j$ and expanding the individual error terms[12] we get $c_i \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + [\,\tilde{\mathbf{d}}_{\pi_1} \quad \tilde{\mathbf{d}}_{\pi_2}\,] \begin{bmatrix} a_i \\ b_i \end{bmatrix} - \tilde{\mathbf{d}}_j = -\mathbf{e}'_j + \mathbf{e}'_{\pi_0} + [\,\mathbf{e}'_{\pi_1} \quad \mathbf{e}_{\pi_2}{}'\,] \begin{bmatrix} a_i \\ b_i \end{bmatrix} - [\,\mathbf{e}'_{\pi_0} \quad \mathbf{e}_{\pi_0}{}'\,] \begin{bmatrix} a_i \\ b_i \end{bmatrix}$. A bound on the uncertainty region for the sum of the errors on the right hand side of this equation is given by the Minkowsky sum of the individual uncertainty regions for $\mathbf{e}'_{\rho_0}$, $\mathbf{e}'_{\rho_1}$, $\mathbf{e}'_{\rho_2}$, and $\mathbf{e}'_j$, which leads to constraints for each match on the two viewpoint parameters, ξ_1 and ξ_2.[2]

$$(c_i \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + [\,\tilde{\mathbf{d}}_{\rho_1} \quad \tilde{\mathbf{d}}_{\rho_2}\,] \begin{bmatrix} a_i \\ b_i \end{bmatrix} - \tilde{\mathbf{d}}_j)^t \hat{\mathbf{n}}_l \geq -\epsilon^l(|a_i + b_i - 1| + |a_i| + |b_i| + 1) \quad (5)$$

The inequality (5) represent conservative constraints on the parameters ξ_1 and ξ_2 feasible for feature match $(\mathbf{m}_i, \tilde{\mathbf{d}}_j)$ when the primary model and data features are approximately aligned; these constraints are based directly on bounds on measurement error for $\mathbf{d}'_{\rho_0}$, $\mathbf{d}'_{\rho_1}$, $\mathbf{d}'_{\rho_2}$, and $\mathbf{d}'_j$.

After this change to affine bases we simply analyze a 2D constraint arrangement for the two components ξ_1 and ξ_2 which is now an arrangement of convex constraint polygons, one for each secondary feature match. Each polygon's shape is a scaled copy of the basic data uncertainty region and its size is roughly proportional to $(|a_i| + |b_i| + 1)/c_i$. This arrangement of MN such polygons is now only of size $O(M^2N^2)$, however we increased the combinatorial component to time $O(M^3N^3)$ because we must search over the possible triples of primary model and data features which form the affine bases. These primary matches must be correctly matched to find geometrically consistent secondary matches. This formulation for deriving constraints on ξ_1 and ξ_2 for each match is closely related to an approach working directly in the 6D space of s-parameters s_{ij} using the constraints of section 3 and making a careful projection of the constraints onto a lower dimensional space[8].

In the special case where the model features are coplanar we have that $c_i = 0$ for all i, and 3D rigid motion corresponds to a 2D affine transformation. If we use the vector $\mathbf{m}_{\pi_1}$ and its perpendicular $\mathbf{m}^{\perp}_{\pi_1}$ as a basis to represent the coplanar model features, the image of the model over all viewpoints is described by

[2] The bound on the feasible viewpoint parameters ξ_1 and ξ_2 is essentially the same as the bounds on the reprojection of a secondary model feature by an alignment transformation computed from three primary matches in the case of *co-planar* model features[12].

$\mathbf{Sm}_i = [\,\mathbf{d}_{\rho_1} \quad \boldsymbol{\eta}\,][\,a_i \quad b_i\,]^t$ where the two parameters $\boldsymbol{\eta} = [\,\eta_1 \quad \eta_2\,]^t$ completely characterized the viewpoint. This leads to constraints on the feasible parameters $\boldsymbol{\eta}$ for a given secondary match $(\mathbf{m}_i, \tilde{\mathbf{d}}_j)$:
$([\,\tilde{\mathbf{d}}_{\rho_1}\boldsymbol{\eta}\,]\begin{bmatrix} a_i \\ b_i \end{bmatrix} - \tilde{\mathbf{d}}_j)^t\hat{\mathbf{n}}_l \geq -\epsilon^l(|a_i - 1| + |a_i| + 1)$. We then have $O(M^2N^2)$ combinatorial component, and a planar constraint analysis in the η_1-η_2 plane as before. See Cass[9] for more detail, and Cass[6, 8] and Hopcroft[14] for related approaches. Similar constraints can be developed for the case of scaled planar motion[7].

4 An Algorithm for Square Uncertainty Bounds

We can now pull together the ideas from the previous sections to develop an algorithm for geometric matching of 3D models to 2D data. A basic algorithm for geometric feature matching consists of performing a combinatorial search through triples of feature matches each defining affine bases for the model and data points. For each such primary match triple, compute for each secondary feature match the constraint polygon for the set of feasible transformation parameters ξ_1 and ξ_2 and analyze the 2D constraint arrangement to find match sets with a high conservative estimate of geometric consistency.

Consider the constraint formulation using axial square uncertainty bounds for the position of the data features characterized by $|\hat{\mathbf{x}}^t\mathbf{e}'_j| \leq \epsilon$ and $|\hat{\mathbf{y}}^t\mathbf{e}'_j| \leq \epsilon$ for all j. Assume we have selected a triple of primary feature matches yielding $(\mathbf{m}_{\pi_1}, \tilde{\mathbf{d}}_{\rho_1})$ and $(\mathbf{m}_{\pi_2}, \tilde{\mathbf{d}}_{\rho_2})$ as primary matches. In this case equation (5) yields
$|(c_i\begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + [\,\tilde{\mathbf{d}}_{\rho_1} \quad \tilde{\mathbf{d}}_{\rho_2}\,]\begin{bmatrix} a_i \\ b_i \end{bmatrix} - \tilde{\mathbf{d}}_j)^t\hat{\mathbf{x}}| \leq \epsilon(|a_i + b_i - 1| + |a_i| + |b_i| + 1)$ and
$|(c_i\begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + [\,\tilde{\mathbf{d}}_{\rho_1} \quad \tilde{\mathbf{d}}_{\rho_2}\,]\begin{bmatrix} a_i \\ b_i \end{bmatrix} - \tilde{\mathbf{d}}_j)^t\hat{\mathbf{y}}| \leq \epsilon(|a_i + b_i - 1| + |a_i| + |b_i| + 1)$ as constaints on the *unknown* vector $[\,\xi_1 \quad \xi_2\,]^t$ feasible with respect to the secondary feature match $(\mathbf{m}_i, \tilde{\mathbf{d}}_j)$. The set of feasible viewpoint parameters $[\,\xi_1 \quad \xi_2\,]^t$ for the feature match $(\mathbf{m}_i, \tilde{\mathbf{d}}_j)$ is bounded by a square of side $\frac{2\epsilon}{c_i}(|a_i + b_i - 1| + |a_i| + |b_i| + 1)$ centered at the point $\frac{1}{c_i}(\tilde{\mathbf{d}}_j - [\,\tilde{\mathbf{d}}_{\rho_1} \quad \tilde{\mathbf{d}}_{\rho_2}\,][\,a_i \quad b_i\,]^t)$ in the ξ_1-ξ_2 plane.

For every secondary match $(\mathbf{m}_i, \tilde{\mathbf{d}}_j)$ we have such a square constraint on the feasible non-linear component $[\,\xi_1 \quad \xi_2\,]^t$. There are $(M-3)(N-3)$ of them forming an arrangement with $O(M^2N^2)$ cells in the ξ_1-ξ_2 plane. Each cell of this arrangement corresponds to a set of values for the parameters $[\,\xi_1 \quad \xi_2\,]^t$ for which a set of feature matches are consistent with respect to these approximate bounds. Finding cells consistent for many feature matches yield match sets and transformations which are very good candidates for large, fully geometrically consistent match sets. An important point, however, is that although there are $O(M^2N^2)$ cells, finding the cells of maximal coverage in this arrangment, e.g. computing $\Phi^{\mathcal{M}}(T) = |\varphi(T)|$ defined in section 1.1 can be done in $O(MN \lg MN)$ time as discussed in section 4.1. It is more difficult to compute the measure $\Phi(T)$, with $T = [\,\xi_1 \quad \xi_2\,]^t$, described in equation (1). We do not know of an

$O(MN \lg MN)$ time algorithm for this, although we use a heuristic algorithm with very good performance.

4.1 Analyzing the Planar Constraint Arrangement

Define the set $\mathcal{F}_\xi^{ij}$ to be the set of parameters (ξ_1, ξ_2) satisfying equation (5) for feature match $(\mathbf{m}_i, \tilde{\mathbf{d}}_j)$; $\mathcal{F}_\xi^{ij}$ is an axial square in the ξ_1-ξ_2 plane. Also define $\varphi_\xi(\xi_1, \xi_2) = \{(\mathbf{m}_i, \tilde{\mathbf{d}}_j) | (\xi_1, \xi_2) \in \mathcal{F}_\xi^{ij}\}$. Then analogous to equation (1) we have $\Phi_\xi(\xi_1, \xi_2)$ and similarly $\Phi_\xi^{\mathcal{M}} = |\varphi_\xi(\xi_1, \xi_2)|$. Computing the match count $\Phi_\xi^{\mathcal{M}}$ is straightforward. Compute the MN axial squares for the matches as described in section 4. We seek a point in the ξ_1-ξ_2 plane of maximal coverage by the squares. This can be computed, for MN axial rectangles in time $O(MN \lg MN)$ using a segment tree[19]. To compute $\Phi_\xi(\xi_1, \xi_2)$ we use a heuristic extention to the same $O(MN \lg MN)$ time algorithm. It is heuristic in that it can only be guaranteed to produce a lower bound to the maximal value of $\Phi_\xi(\xi_1, \xi_2)$, but performs very well empirically.

4.2 A Tradeoff: Geometric Consistency vs. Algorithmic Complexity

There is an important tradeoff we made in the previous section between finding fully geometrically consistent match sets, and the nature of the algorithm required to do so. In effect we moved part of the computation away from the complex computational geometric problem of analyzing constraint arrangements in higher dimensional spaces, to a simple combinatorial search through triples of primary matches, leaving the much smaller and simpler computational geometric problem of analyzing planar arrangements. Analyzing arrangements is hard algorithmically, combinatorial searches are easy; so this is a good trade. This also allows us to incorporate other methods for reducing the computation required like grouping of primary data triples. For this reduction we traded somewhat relaxed geometric consistency constraints. It is, however, a careful approximation and a *conservative estimate* of full geometric consistency; we will always overestimate the consistency.

5 Testing the Method

The method was tested on both synthetic data and real image data, both for the cases of 2D object undergoing 3D motion using the η-plane variation, and 3D objects undergoing 3D motion using the ξ-plane variation. Here we present a representative sample of tests for 3D objects in 3D. The size of the secondary feature match constraint squares in the ξ-plane was computed using a Gaussian approximation method[3] instead of the absolute bound as discussed in section 4 because the performance was better.

[3] Instead of the uniform distribution of $\mathbf{e}'_{\rho_0}$, $\mathbf{e}'_{\rho_1}$, and $\mathbf{e}'_{\rho_2}$ assume they are indentically distributed Gaussian random variables with variance $\sigma_e^2 = \epsilon^2/4$. Our bounded regions

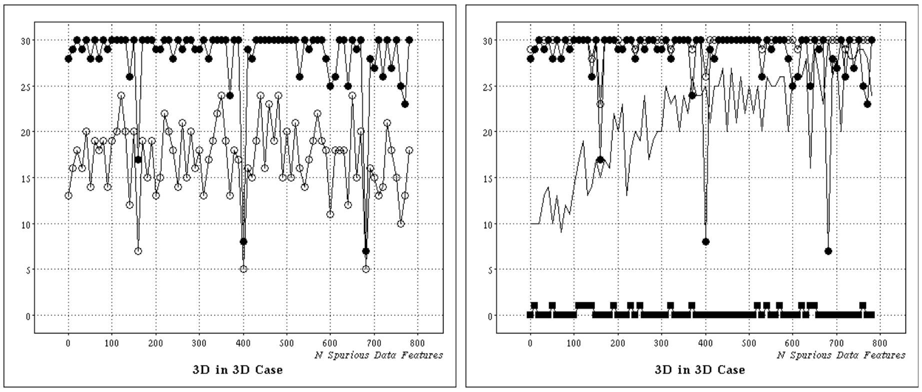

Fig. 1. Results from synthetic tests. In the left graph, the curve with solid dots shows the number correctly matched features aligned in the image within 2ϵ, the curve with open dots shows the number correctly matched features aligned in the image within ϵ. The value plotted is the maximum, over three different correctly chosen primary match groups. In the right graph, the open dots show correct matches aligned within 2ϵ, the filled dots show the consistency measure for matches aligned within 2ϵ (the two curves are almost coincident) in both cases for correct primary matches. The curve labelled with squares (at the bottom) shows the number of correct matches aligned within 2ϵ, and the plain curve shows the consistency measure for matches aligned within 2ϵ, in both these latter graphs for the case of incorrect primary matches. Here $M = 30$, $\epsilon = 5$, the image size was 1024^2, and $N = 1,800$.

The synthetic model objects consisted of points generated uniformly in in a sphere in the 3D in 3D case, of about the size of the image dimensions. The models were imaged synthetically by applying a random transformation and projection to the image plane, then adding independent error to each point uniformly distributed over a square of side 2ϵ. Spurious data features were then generated uniformly in a circle the size of the image. The experimental results show that the method works well for finding transformations with high geometric consistency, as shown in figure 1.

Tests of the method have also been performed on data derived from real images of both 2D and 3D objects. For a 3D object, two black metal bookends where fastened together to construct an interesting non-planar object. To speed up the experiments, a rough grouping was done by hand so as to avoid the long combinatorial search through primary matches, though no specific primary matches were specified. See figure 2.

---

enclose approximately distances within 2σ. The vector $\boldsymbol{\xi}$ is then a linear function of these random variables and is thus also a Gaussian random variable. The variance of $\boldsymbol{\xi}$ is then $\sigma^2_{\xi_i} = \frac{4\epsilon^2}{c_i^2}[(a_i + b_i - 1)^2 + a_i^2 + b_i^2 + 1)]$.

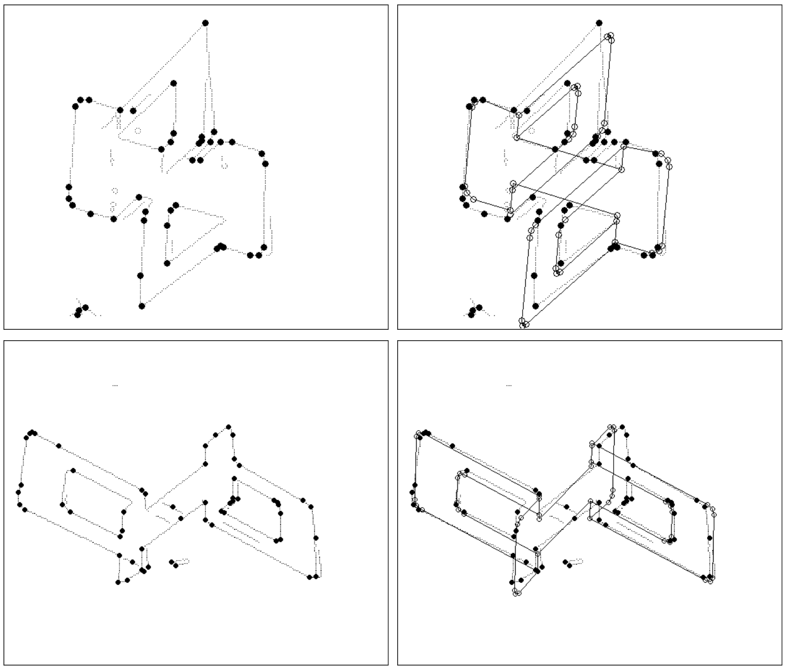

Fig. 2. Experments with real images. On the left are the image features, connected by line segments. On the right is the best model transformation, shown by overlaying the model. Note that no verification has been performed—shown are the raw hypotheses computed assuming 3D affine transformations.

6 Conclusion

We briefly note some related work. The most similar geometric matching method to ours is the alignment method[16] in it's more careful form accounting for the effects of data feature error [1]. Our approach does not enforce the quadratic constraints on $\boldsymbol{\xi}$ while alignment does; however our approach requires all secondary matches to use the same $\boldsymbol{\xi}$, while the alignment method does not. For 3D objects, the ξ-plane method, for $M \leq N$, takes $O(M^4 N^4 \lg N)$ time, while the alignment method takes $O(M^4 N^3 \lg N)$. For 2D objects the η-plane method takes time $O(M^3 N^3 \lg N)$ and alignment again takes $O(M^4 N^3 \lg N)$, yet by randomly choosing primary matches, they both take $O(N^3 M \lg N)$ expected time.

We note there are other approaches based on ideas from computational geometry and grounded in real recognition systems focused on 2D objects and affine transformations[15, 14, 20].

The author thanks Tao Alter, David Jacobs, Dan Huttenlocher, and William Rucklidge for helpful comments and discussions.

References

1. Tao D. Alter. Robust and efficient 3d recognition by alignment. Master's thesis, Massachusetts Institute of Technology, 1992.
2. Tao D. Alter. 3-d pose from 3 points using weak-perspective. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(8):802–808, 1994.
3. Henry S. Baird. *Model-Based Image Matching Using Location.* MIT Press, Cambridge, MA, 1985.
4. Thomas M. Breuel. An efficient correspondence based algorithm for 2d and 3d model based recognition. A.I. Memo No. 1259, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, October 1990.
5. Todd A. Cass. Parallel computation in model-based recognition. Master's thesis, Massachusetts Institute of Technology, 1988.
6. Todd A. Cass. *Polynomial-Time Geometric Matching for Object Recognition.* PhD thesis, Massachusetts Institute of Technology, 1992.
7. Todd A. Cass. Polynomial-time geometric matching for object recognition. In *Proceedings of the European Conference on Computer Vision*, 1992.
8. Todd A. Cass. Robust geometric matching for 3d object recognition. In *Proceedings Int. Conf. on Pattern Recognition*, pages A477–A482, Jerusalem, Israel, October 1994.
9. Todd A. Cass. Robust affine structure matching for 3d object recognition. *Submitted for publication review*, Oct 1995.
10. Todd A. Cass. The three-point pose solution under scaled-orthography. *Submitted for publication review*, Oct 1995.
11. Herbert Edelsbrunner. *Algorithms in Combinatorial Geometry.* Springer-Verlag, 1987.
12. W. Eric L. Grimson, Daniel P. Huttenlocher, and David W. Jacobs. Affine matching with bounded sensor error: A study of geometric hashing & alignment. Technical Report A.I. Memo 1250, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1990.
13. W. Eric L. Grimson and Tomas Lozano-Perez. Localizing overlapping parts by searching the interpretation tree. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(4):469–482, July 1987.
14. Michael J. Hopcroft. *A Geometrical Approach to Model-Based Vision.* PhD thesis, Cornell University, 1995.
15. D. Huttenlocher, G. Klanderman, and W. Rucklidge. Comparing images using the hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9):850–863, 1993.
16. D. Huttenlocher and S. Ullman. Recognizing solid objects by alignment with an image. *International Journal of Computer Vision*, 5(7):195–212, 1990.
17. Daniel P. Huttenlocher and Todd A. Cass. Measuring the quality of hypotheses in model-based recognition. In *Proceedings of the European Conference on Computer Vision*, Genova, Italy, May 1992.
18. David W. Jacobs. 2d images of 3d oriented points. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, New York, NY, June 1993.
19. E. P. Preparata and M. Shamos. *Computational Geometry–An Introduction.* Springer-Verlag, New York, 1985.
20. William Rucklidge. Locating objects using the hausdorff distance. In *Proceedings of the International Conference on Computer Vision*, Cambridge, MA,, June 1995.