A Study of Affine Matching With Bounded Sensor Error *

W. Eric L. Grimson[1], *Daniel P. Huttenlocher*[2] *and David W. Jacobs*[1]

[1] AI Lab, Massachusetts Institute of Technology, Cambridge MA 02139, USA
[2] Department of Computer Science, Cornell University, Ithaca NY 14853, USA

Abstract. Affine transformations of the plane have been used by model-based recognition systems to approximate the effects of perspective projection. Because the underlying mathematics are based on *exact* data, in practice various heuristics are used to adapt the methods to real data where there is positional uncertainty. This paper provides a precise analysis of affine point matching under uncertainty. We obtain an expression for the range of affine-invariant values consistent with a given set of four points, where each data point lies in an ϵ-disc. This range is shown to depend on the actual x-y-positions of the data points. Thus given uncertainty in the data, the representation is no longer invariant with respect to the Cartesian coordinate system. This is problematic for methods, such as geometric hashing, that depend on the invariant properties of the representation. We also analyze the effect that uncertainty has on the probability that recognition methods using affine transformations will find false positive matches. We find that such methods will produce false positives with even moderate levels of sensor error.

1 Introduction

In the model-based approach to object recognition, a set of geometric features from an object model are compared against like features from an image of a scene (cf. [3, 9]). This comparison generally involves determining a valid correspondence between a subset of the model features and a subset of the image features, where valid means there exists some transformation of a given type mapping each model feature onto its corresponding image feature. The quality of an hypothesized transformation is then evaluated by determining if the number of features brought into correspondence accounts for a sufficiently large portion of the model and the data.

Several recent systems have used affine transformations of the plane to represent the mapping from a 2D model to a 2D image (e.g. [4, 5, 15, 16, 20, 21, 22, 24, 25, 28]). This type of transformation also approximates the 2D image of a planar object at an arbitrary orientation in 3D space, and is equivalent to a 3D rigid motion of the object, followed by orthographic projection and scaling (dilation). The scale factor accounts for the perceptual shrinking of objects with distance. This affine viewing model does not

* This report describes research done in part at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's research is provided in part by an ONR URI grant under contract N00014-86-K-0685, and in part by DARPA under Army contract number DACA76-85-C-0010 and under ONR contract N00014-85-K-0124. WELG is supported in part by NSF contract number IRI-8900267. DPH is supported at Cornell University in part by NSF grant IRI-9057928 and matching funds from General Electric and Kodak, and in part by AFOSR under contract AFOSR-91-0328.

capture the perspective distortions of real cameras, but it is a reasonable approximation to perspective except when an object is deep with respect to its distance from the viewer.

Recognition systems that use 2D affine transformations fall into two basic classes. Methods in the first class explicitly compute an affine transformation based on the correspondence of a set of 'basis features' in the image and the model. This transformation is applied to the remaining model features, mapping them into the image coordinate frame where they are compared with image features [2, 15, 16, 24]. Methods in the second class compute and directly compare affine invariant representations of the model and the image [4, 5, 20, 21, 22, 25, 28] (there is also recent work on deriving descriptions of shapes that are invariant under perspective projection [8, 27]). In either case, recognition systems that employ affine transformations generally use some heuristic means to allow for uncertainty in the location of sensory data. One notable exception is [4] who formulate a probabilistic method. [26] also discusses bounds on the effects of error on invariants, and [7] addresses this problem for simpler similarity transformations. In previous work [17] we provided a precise account of how uncertainty in the image measurements affects the range of transformations consistent with a given configuration of points acting under an affine transformation. Here, we show that many existing recognition methods are not actually able to find instances of an object, without also admitting a large number of false matches. The analysis further suggests techniques for developing new recognition methods that will explicitly account for uncertainty.

1.1 Affine Transformations and Invariant Representations

An affine transformation of the plane can be represented as a nonsingular 2×2 matrix $\mathbf{L}$, and a 2-vector, $\mathbf{t}$, such that a given point $\mathbf{x}$ is transformed to $\mathbf{x}' = \mathbf{L}\mathbf{x} + \mathbf{t}$. Such a transformation can be defined to map any triple of points to any other triple (expect in degenerate cases). As well, three points define an affine coordinate frame (analogous to a Cartesian coordinate frame in the case of Euclidean transformations) [6, 18], e.g., given a set of points $\{\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3\}$, any other point $\mathbf{x}$ can be expressed as:

$$\mathbf{x} = \mathbf{m}_1 + \alpha(\mathbf{m}_2 - \mathbf{m}_1) + \beta(\mathbf{m}_3 - \mathbf{m}_1). \tag{1}$$

α and β remain unchanged when any affine transformation A is applied to the points:

$$A(\mathbf{x}) = A(\mathbf{m}_1) + \alpha(A(\mathbf{m}_2) - A(\mathbf{m}_1)) + \beta(A(\mathbf{m}_3) - A(\mathbf{m}_1)).$$

Thus the pair (α, β) constitute affine-invariant coordinates of $\mathbf{x}$ with respect to the *basis* $(\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3)$. We can think of (α, β) as a point in a 2D space, termed the α-β-plane.

The main issue we wish to explore is: Given a model basis of three points and some other model point, what sets of four image features are possible transformed instances of these points? The exact location of each image feature is unknown, and thus we model image features as discs of radius ϵ. The key question is what effect this uncertainty has on which image quadruples are possible transformed instances of a model quadruple.

We assume that a set of model points is given in a Cartesian coordinate frame, and some distinguished basis triple is also specified. Similarly a set of image points is given in their coordinate frame. Two methods can be used to map between the model and the image. One method, used by geometric hashing [20], maps both model and image points to (α, β) values using the basis triples. The other method, used by alignment [15], computes the transformation mapping the model basis to the image basis, and uses it to map all model points to image coordinates. In both cases, a distinguished set of three

model and image points is used to map a fourth point (or many such points) into some other space. We consider the effects of uncertainty on these two methods.

First we characterize the range of image measurements in the x-y (Euclidean) plane that are consistent with the (α, β) pair computed for a given quadruple of model points, as specified by equation (1). This corresponds to explicitly computing a transformation from one Cartesian coordinate frame (the model) to another (the image). We find that if sensor points' locational uncertainty is bounded by a disc of radius ϵ, then the range of possible image measures consistent with a given (α, β) pair is a disc with radius between $\epsilon(1+|\alpha|+|\beta|)$ and $2\epsilon(1+|\alpha|+|\beta|)$. This defines the set of image points that could match a specific model point, given both an image and model basis.

We then perform the same analysis for the range of affine coordinate, (α, β), values that are consistent with a given quadruple of points. This corresponds to mapping both the model and image points to (α, β) values. To do this, we use the expressions derived for the Euclidean case to show that the region of α-β-space that is consistent with a given point and basis, is in general an ellipse containing the point (α, β). The parameters of the ellipse depend on the actual locations of the points defining the basis. Hence the set of possible values in the α-β-plane *cannot* be computed independent of the actual locations of the image basis points. In other words there is an interaction between the uncertainty in the sensor values and the actual locations of the sensor points. This limits the applicability of methods which assume that these are independent of one another. For example, the geometric hashing method requires that the α-β coordinates be independent of the actual location of the basis points in order to construct a hash table.

2 Image Uncertainty and Affine Coordinates

Consider a set of three model points, $\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3$, and the affine coordinates (α, β) of a fourth model point $\mathbf{x}$ defined by

$$\mathbf{x} = \mathbf{m}_1 + \alpha(\mathbf{m}_2 - \mathbf{m}_1) + \beta(\mathbf{m}_3 - \mathbf{m}_1) \tag{2}$$

plus a set of three sensor points $\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3$, such that $\mathbf{s}_i = T(\mathbf{m}_i) + \mathbf{e}_i$, where T is some affine transformation, and $\mathbf{e}_i$ is an arbitrary vector of magnitude at most ϵ_i. That is, T is some underlying affine transformation that cannot be directly observed in the data because each data point is known only to within a disc of radius ϵ_i.

We are interested in the possible locations of a fourth sensor point, call it $\hat{\mathbf{x}}$, such that $\hat{\mathbf{x}}$ could correspond to the ideally transformed point $T(\mathbf{x})$. The possible positions of $\hat{\mathbf{x}}$ are affected both by the error in measuring each image basis point, $\mathbf{s}_i$, and by the error in measuring the fourth point itself. Thus the possible locations are given by transforming equation (2) and adding in the error $\mathbf{e}_0$ from measuring $\mathbf{x}$,

$$\begin{aligned}\hat{\mathbf{x}} &= T\left(\mathbf{m}_1 + \alpha(\mathbf{m}_2 - \mathbf{m}_1) + \beta(\mathbf{m}_3 - \mathbf{m}_1)\right) + \mathbf{e}_0 \\ &= \mathbf{s}_1 + \alpha(\mathbf{s}_2 - \mathbf{s}_1) + \beta(\mathbf{s}_3 - \mathbf{s}_1) - \mathbf{e}_1 + \alpha(\mathbf{e}_1 - \mathbf{e}_2) + \beta(\mathbf{e}_1 - \mathbf{e}_3) + \mathbf{e}_0.\end{aligned}$$

The measured point $\hat{\mathbf{x}}$ can lie in a range of locations about the ideal location $\mathbf{s}_1 + \alpha(\mathbf{s}_2 - \mathbf{s}_1) + \beta(\mathbf{s}_3 - \mathbf{s}_1)$ with deviation given by the linear combination of the four error vectors:

$$-\mathbf{e}_1 + \alpha(\mathbf{e}_1 - \mathbf{e}_2) + \beta(\mathbf{e}_1 - \mathbf{e}_3) + \mathbf{e}_0 = -[(1 - \alpha - \beta)\mathbf{e}_1 + \alpha\mathbf{e}_2 + \beta\mathbf{e}_3 - \mathbf{e}_0]. \tag{3}$$

The set of possible locations specified by a given $\mathbf{e}_i$ is a disc of radius ϵ_i about the origin:

$$C(\epsilon_i) = \{\mathbf{e}_i \mid \|\mathbf{e}_i\| \leq \epsilon_i\}.$$

Similarly, the product of any constant k with $\mathbf{e}_i$ yields a disc $C(k\epsilon_i)$ of radius $|k|\epsilon_i$ centered about the origin. Thus substituting the expressions for the disc in equation (3), the set of all locations about the ideal point $\mathbf{s}_1 + \alpha(\mathbf{s}_2 - \mathbf{s}_1) + \beta(\mathbf{s}_3 - \mathbf{s}_1)$ is:

$$C([1 - \alpha - \beta]\epsilon_1) \oplus C(\alpha\epsilon_2) \oplus C(\beta\epsilon_3) \ominus C(\epsilon_0), \tag{4}$$

where $\oplus$ is the Minkowski sum, i.e. $A \oplus B = \{p + q | p \in A, q \in B\}$ (similarly for $\ominus$).

In order to simplify the expression for the range of $\hat{\mathbf{x}}$ we make use of the following fact, which follows directly from the definition of the Minkowski sum for sets.

Claim 1 *$C(r_1) \oplus C(r_2) = C(r_1) \ominus C(r_2) = C(r_1 + r_2)$, where $C(r_i)$ is a disc of radius r_i centered about the origin, $r_i \geq 0$.*

If we assume that the $\epsilon_i = \epsilon, \forall i$, then Claim 1 simplifies equation (4) to

$$C\left(\epsilon[|1 - \alpha - \beta| + |\alpha| + |\beta| + 1]\right).$$

The absolute values arise because α and β can become negative, but the radius of a disc is a positive quantity. Clearly the radius of the error disc grows with increasing magnitude of α and β, but the actual expression governing this growth is different for different portions of the $\alpha - \beta$-plane, as shown in figure 1.

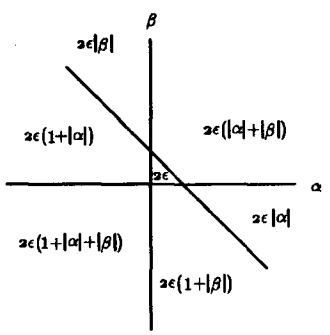

Fig. 1. Diagram of error effects. The region of feasible points is a disc, whose radius is given by the indicated expression, depending on the values of α and β. The diagonal line is $1 - \alpha - \beta = 0$.

We can bound the expressions defining the radius of the uncertainty disc by noting:

$$1 + |\alpha| + |\beta| \leq (|1 - \alpha - \beta| + |\alpha| + |\beta| + 1) \leq 2(1 + |\alpha| + |\beta|).$$

We have thus established the following result, illustrated in figure 2:

Proposition 1. *The range of image locations that is consistent with a given pair of affine coordinates (α, β) is a disc of radius r, where*

$$\epsilon(1 + |\alpha| + |\beta|) \leq r \leq 2\epsilon(1 + |\alpha| + |\beta|)$$

and where $\epsilon \geq 0$ is a constant bounding the positional uncertainty of the image data.

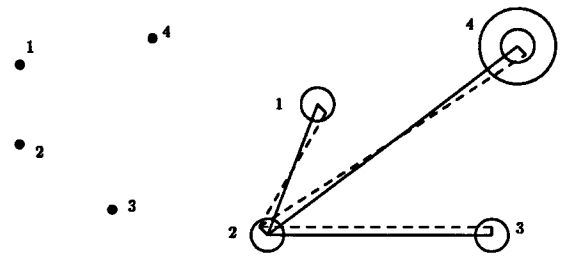

Fig. 2. Diagram of error effects. On the left are four model points, on the right are four image points, three of which are used to establish a basis. The actual position of each transformed model point corresponding to the basis image points is offset by an error vector of bounded magnitude. The coordinates of the fourth point, written in terms of the basis vectors, can thus vary from the ideal case, shown in solid lines, to cases such as that shown in dashed lines. This leads to a disc of variable size in which the corresponding fourth model point could lie.

The expression in Proposition 1 allows the calculation of error bounds for any method based on 2D affine transformations, such as [2, 15, 24]. In particular, if $|\alpha|$ and $|\beta|$ are both less than 1, then the error in the position of a point is at most 6ϵ. This condition can be met by using as the affine basis, three points $\mathbf{m}_1, \mathbf{m}_2$ and $\mathbf{m}_3$ that lie on the convex hull of the set of model points, and are maximally separated from one another.

The expression is independent of the actual locations of the model or image points, so that the possible positions of the fourth point vary only with the sensor error and the values of α and β. They do not vary with the configuration of the model basis (e.g., even if close to collinear) nor do they vary with the configuration of the image basis. Thus, the error range does not depend on the viewing direction. Even if the model is viewed end on, so that all three model points appear nearly collinear, or if the model is viewed at a small scale, so that all three model points are close together, the size of the region of possible locations of the fourth model point in the image will remain unchanged.

The viewing direction does, however, greatly affect the affine coordinate system defined by the three projected model points. Thus the set of possible *affine coordinates* of the fourth point, when considered directly in α-β-space, will vary greatly. Proposition 1 defines the set of image locations consistent with a fourth point. This implicitly defines the set of affine transformations that produce possible fourth image point locations, which can be used to characterize the range of (α, β) values consistent with a set of four points.

We will do the analysis using the upper bound on the radius of the error disc from Proposition 1. In actuality, the analysis is slightly more complicated, because the expression governing the disc radius varies as shown in figure 1. For our purposes, however, considering the extreme case is sufficient. It should also be noted from the figure that the extreme case is in fact quite close to the actual value over much of the range of α and β.

Given a triple of image points that form a basis, and a fourth image point, $\mathbf{s}_4$, we want the range of affine coordinates for the fourth point that are consistent with the possibly erroneous image measurements. In effect, each sensor point $\mathbf{s}_i$ takes on a range of possible values, and each quadruple of such values produces a possibly distinct value using equation (1). As illustrated in figure 3 we could determine all the feasible values by varying the basis vectors over the uncertainty discs associated with their endpoints, finding the set of (α', β') values such that the resulting point in this affine basis lies within ϵ of the original point. By our previous results, however, it is equivalent to find affine coordinates (α', β') such the Euclidean distance from $\mathbf{s}_1 + \alpha'(\mathbf{s}_2 - \mathbf{s}_1) + \beta'(\mathbf{s}_3 - \mathbf{s}_1)$

to $\mathbf{s}_1 + \alpha(\mathbf{s}_2 - \mathbf{s}_1) + \beta(\mathbf{s}_3 - \mathbf{s}_1)$ is bounded above by $2\epsilon(1 + |\alpha'| + |\beta'|)$.

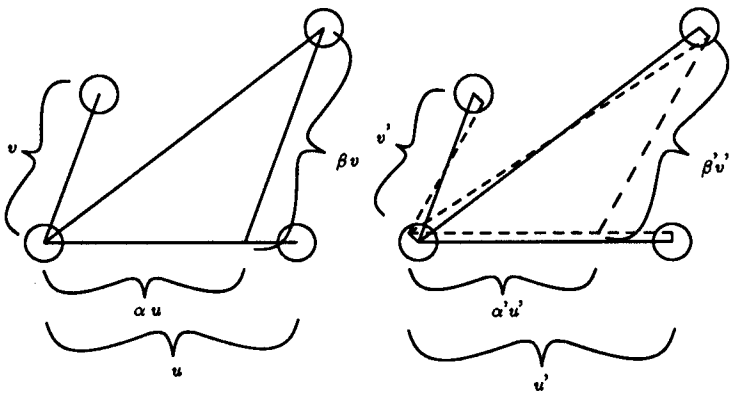

Fig. 3. On the left is a canonical example of affine coordinates. The fourth point is offset from the origin by a scaled sum of basis vectors, $\alpha\mathbf{u} + \beta\mathbf{v}$. On the right is a second consistent set of affine coordinates. By taking other vectors that lie within the uncertainty regions of each image point, we can find different sets of affine coordinates α', β' such that the new fourth point based on these coordinates also lies within the uncertainty bound of the image point.

The boundary of the region of such points (α', β') occurs when the distance from the nominal image point $\mathbf{s}_4 = \mathbf{s}_1 + \alpha(\mathbf{s}_2 - \mathbf{s}_1) + \beta(\mathbf{s}_3 - \mathbf{s}_1)$ is $2\epsilon(1 + |\alpha'| + |\beta'|)$, i.e.

$$[2\epsilon(1 + |\alpha'| + |\beta'|)]^2 = [(\alpha - \alpha')u]^2 + 2(\beta - \beta')(\alpha - \alpha')vu\cos\phi + [(\beta - \beta')v]^2 \quad (5)$$

where $\mathbf{u} = \mathbf{s}_2 - \mathbf{s}_1, \mathbf{v} = \mathbf{s}_3 - \mathbf{s}_1, u = \|\mathbf{u}\|, v = \|\mathbf{v}\|$ and where the angle made by the image basis vectors $\mathbf{s}_2 - \mathbf{s}_1$ and $\mathbf{s}_3 - \mathbf{s}_1$ is ϕ. Considered as an implicit function of α', β', equation (5) defines a conic. If we expand out equation (5), we get

$$a_{11}(\alpha')^2 + 2a_{12}\alpha'\beta' + a_{22}(\beta')^2 + 2a_{13}\alpha' + 2a_{23}\beta' + a_{33} = 0 \quad (6)$$

where

$$a_{11} = u^2 - 4\epsilon^2 \qquad a_{22} = v^2 - 4\epsilon^2$$
$$a_{12} = vu\cos\phi - 4s_\alpha s_\beta\epsilon^2 \qquad a_{13} = -u[\alpha u + \beta v\cos\phi] - 4s_\alpha\epsilon^2$$
$$a_{23} = -v[\alpha u\cos\phi + \beta v] - 4s_\beta\epsilon^2 \quad a_{33} = \alpha^2u^2 + 2\alpha\beta uv\cos\phi + \beta^2v^2 - 4\epsilon^2$$

and where s_α denotes the sign of α, with $s_0 = 1$.

We can use this form to compute the invariant characteristics of a conic [19]:

$$I = u^2 + v^2 - 8\epsilon^2 \quad (7)$$
$$D = u^2v^2\sin^2\phi - 4\epsilon^2\left(u^2 - 2uvs_\alpha s_\beta\cos\phi + v^2\right) \quad (8)$$
$$A = -4\epsilon^2u^2v^2\sin^2\phi(1 + s_\alpha\alpha + s_\beta\beta)^2 \quad (9)$$

If $u^2 + v^2 > 8\epsilon^2$, then $\frac{A}{I} < 0$. Furthermore, if

$$u^2v^2\sin^2\phi > 4\epsilon^2\left(u^2 - 2uvs_\alpha s_\beta\cos\phi + v^2\right)$$

then $D > 0$ and the conic defined by equation (5) is an ellipse. These conditions are not met only when the image basis points are very close together, or when the image basis points are nearly collinear. For instance, if the image basis vectors $\mathbf{u}$ and $\mathbf{v}$ are each at least 2ϵ in length then $u^2 + v^2 > 8\epsilon^2$. Similarly, if $\sin\phi$ is not small, $D > 0$. In fact, cases where these conditions do not hold will be very unstable and should be avoided.

We can now compute characteristics of the ellipse. The area of the ellipse is given by

$$\frac{4\pi\epsilon^2 u^2 v^2 \sin^2\phi(1 + s_\alpha\alpha + s_\beta\beta)^2}{[u^2v^2\sin^4\phi - 4\epsilon^2(u^2 - 2uvs_\alpha s_\beta\cos\phi + v^2)]^{\frac{3}{2}}}. \tag{10}$$

The center of the ellipse is at

$$\begin{aligned}\alpha_0 &= D^{-1}\left[\alpha u^2v^2\sin^2\phi - 4\epsilon^2(\alpha u^2 - s_\alpha(1 + s_\beta\beta)v^2 + uv\cos\phi(\beta + s_\beta(1 - s_\alpha\alpha)))\right]\\ \beta_0 &= D^{-1}\left[\beta u^2v^2\sin^2\phi - 4\epsilon^2(\beta v^2 - s_\beta(1 + s_\alpha\alpha)u^2 + uv\cos\phi(\alpha + s_\alpha(1 - s_\beta\beta)))\right].\end{aligned} \tag{11}$$

The angle Φ of the principal axes with respect to the α axis is

$$\tan 2\Phi = \frac{2[uv\cos\phi - 4\epsilon^2 s_\alpha s_\beta]}{u^2 - v^2}. \tag{12}$$

Thus we have established the following:

Proposition 2. *Given bounded errors of ϵ in the measurement of the image points, the region of uncertainty associated with a pair of affine coordinates (α, β) in α-β-space is an ellipse. The area of this ellipse is given by equation (10), the center is at (α_0, β_0) as given by equation (11), and the orientation is given by equation (12).*

Hence, given four points whose locations are only known to within ϵ-discs, there is an elliptical region of possible (α, β) values specifying the location of one point with respect to the other three. Thus if we compare (α, β) values generated by some object model with those specified by an ϵ-uncertain image, each image datum actually specifies an ellipse of (α, β) values, whose area depends on ϵ, α, β, and the configuration of the three image points that form the basis. To compare the model values with image values one must see if the affine-invariant coordinates for each model point lie within the elliptical region of possible affine-invariant values associated with the corresponding image point.

The elliptical regions of consistent parameters in α-β-space cause some difficulties for discrete hashing schemes. For example, geometric hashing uses affine coordinates of model points, computed with respect to some choice of basis, as the hash keys to store the basis in a table. In general, the implementations of this method use square buckets to tessellate the hash space (the α-β-space). Even if we chose buckets whose size is commensurate with the ellipse, several such buckets are likely to intersect any given ellipse due to the difference in shape. Thus, one must hash to multiple buckets, which increases the probability that a random pairing of model and image bases will receive a large number of votes.

A further problem for discrete hashing schemes is that the size of the ellipse increases as a function of $(1 + |\alpha| + |\beta|)^2$. Thus points with larger affine coordinates give rise to larger ellipses. Either one must hash a given value to many buckets, or one must account for this effect by sampling the space in a manner that varies with $(1 + |\alpha| + |\beta|)^2$.

The most critical issue for discrete hashing schemes, such as geometric hashing, is that the shape, orientation and position of the ellipse depend on the specific image basis chosen. Because the error ellipse associated with a given (α, β) pair depends on the

characteristics of the image basis, which are not known until run time, there is no way to pre-compute the error regions and thus no clear way to fill the hash table as a pre-processing step, independent of a given image. It is thus either necessary to approximate the ellipses by assuming bounds on the possible image basis, which will allow both false positive and false negative hits in the hash table, or to compute the ellipse to access at run time. Note that the geometric hashing method does not address these issues. It simply assumes that some 'appropriate' tessellation of the image space exists.

In summary, in this section we have characterized the range of image coordinates and the range of (α, β) values that are consistent with a given point, with respect to some basis, when there is uncertainty in the image data. In the following section we analyze what fraction of all possible points (in some bounded image region) are consistent with a given range of (α, β) values. This can then be used to estimate the probability of a false match for various recognition methods that employ affine transformations.

3 The Selectivity of Affine-Invariant Representations

What is the probability than an object recognition system will erroneously report an instance of an object in an image? Recall that such an instance in general is specified by giving a transformation from model coordinates to image coordinates, and a measure of 'quality' based on the number of model features that are paired with image features under this transformation. Thus we are interested in whether a random association of model and image features can occur in sufficient number to masquerade as a correct solution. We use the results developed above to determine the probability of such a *false match*. There are two stages to this analysis; the first is a statistical analysis that is independent of the given recognition method, and the second is a combinatorial analysis that depends on the particular recognition method. In this section we examine the first stage. In the following section we apply the analysis to the alignment method.

To determine the probability that a match will be falsely reported we need to know the 'selectivity' of a quadruple of model points. Recall that each model point is mapped to a point in α-β-space, with respect to a particular model basis. Similarly each image point, modeled as a disc, is mapped to an elliptical region of possible points in α-β-space. Each such region that contains one or more model points specifies an image point that is consistent with the given model. Thus we need to estimate the probability that a given image basis and fourth image point chosen at random will map to a region of α-β-space that is consistent with one of the model points written in terms of some model basis. This is characterized by the proportion of α-β-space consistent with a given basis and fourth point (where the size of the space is bounded in some way). As shown above, the elliptical regions in α-β-space are equivalent to circular regions in image space. Thus, for ease of analysis we use the formulation in terms of circles in image space.

To determine the selectivity, assume we are given some image basis and a potential corresponding model basis. Each of the remaining $m-3$ model points are defined as affine coordinates relative to the model basis. These can then be transformed into the image domain, by using the same affine coordinates, with respect to the image basis. Because of the uncertainty of the image points, there is an uncertainty in the associated affine transformation. This manifests itself as a range of possible positions for the model points, as they are transformed into the image. Previously we determined that a transformed model point had to be within $2\epsilon(1 + |\alpha| + |\beta|)$ of an image point in order to match it. That calculation took into account error in the matched image point as well as the basis image points. Therefore, placing an appropriately sized disc about each model point is

equivalent to placing an ϵ sized disc about each image point. We thus represent each transformed model point as giving rise to a disc of some radius, positioned relative to the nominal position of the model point with respect to the image basis. For convenience, we use the upper bound on the size of the radius, $2\epsilon(1 + |\alpha| + |\beta|)$. For each model point, we need the probability that at least one image point lies in the associated error disc about the model point transformed to the image, because if this happens then there is a consistent model and image point for the given model and image basis. To estimate this probability, we need the expected size of the disc. Since the disc size varies with $|\alpha| + |\beta|$, this means we need an estimate of the distribution of points with respect to affine coordinates. By figure 1 we should find the distribution of points as a function of (α, β). This is messy, and thus we use an approximation instead.

For this approximation, we measure the distribution with respect to $\rho = |\alpha| + |\beta|$, since both the upper and lower bounds on the disc size are functions of ρ. Intuitively we expect the distribution to vary inversely with ρ. To verify this, we ran the following experiment. A set of 25 points were generated at random, such that their pairwise separation was between 25 and 250 pixels. All possible bases were selected, and for each basis for which the angle between the axes was at least $\pi/16$, all the other model points were rewritten in terms of affine invariant coordinates (α, β). This gave roughly 300,000 samples, which we histogrammed with respect to ρ. We found that the maximum value for ρ in this case was roughly 51. In general, however, almost all of the values were much smaller, and indeed, the distribution showed a strong inverse drop off (see figure (4)). Thus, we use the following distribution of points in affine coordinates:

$$\delta(\alpha, \beta) = \begin{cases} k\rho & \rho \leq 1 \\ k\rho^{-2} & \rho \geq 1. \end{cases} \tag{13}$$

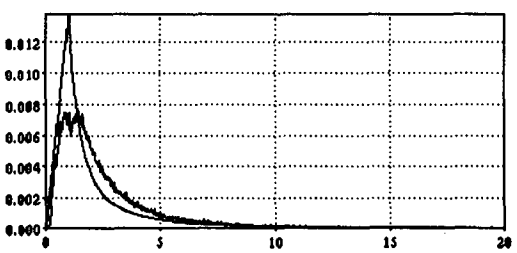

Fig. 4. Histogram of distribution of $|\alpha| + |\beta|$ values. Vertical axis is ratio of number of samples to total samples, horizontal axis is value for $|\alpha| + |\beta|$. The maximum over 300,000 samples was 51. Only the first portion of the graph is displayed. Overlayed with this is the distribution given in equation (13).

Note that this model underestimates the probability for large values of ρ, while overestimating it for small values of ρ. Since we want the expected size of the error disc, and this grows with ρ, such an approximation will underestimate the size of the disc.

First, we integrate equation (13) and normalize to 1 to deduce the constant:

$$k = \frac{2}{3 - 2\rho_m^{-1}} \tag{14}$$

where ρ_m is the maximum value for ρ.

Next, we want to find the expected area of a disc in image space. Recall that we are using the upper bound on disc size, so in principle this area is just $4\pi\epsilon^2(1+\rho)^2$. We could simply integrate this with respect to the distribution from equation (13). This, however, ignores the fact that the image is of finite size (say each dimension is $2r$), and some of the disc may lie beyond the bounds of the image. We therefore separate out four cases.

The first case is for $\rho \leq 1$. Here we get an expected area

$$A_1 = \int_{\rho=0}^{1} 4\pi\epsilon^2(1+\rho)^2 k\rho\, d\rho = 4\pi\epsilon^2 k\frac{17}{12}. \tag{15}$$

The second case considers discs that lie entirely within the image. For convenience, assume that the coordinate frame of the basis is centered at the image center (since the circle is entirely inside the image), and the image dimensions are $2r$ by $2r$. In this case, we have $r - p \geq \gamma$ where $\gamma = 2\epsilon(1+\rho)$. In general, we have $p \leq \rho d$ where d is the separation between two of the basis points in the image, and thus if $1 \leq \rho \leq c_1$ where

$$c_1 = \min\left\{\rho_m, \frac{r-2\epsilon}{2\epsilon+d}\right\}$$

then the discs will all lie entirely within the image. Thus the second case is

$$\begin{aligned} A_2 &= \int_{\rho=1}^{c_1} 4\pi\epsilon^2(1+\rho)^2 k\rho^{-2}\, d\rho = 4\pi\epsilon^2 k\left[c_1 + 2\log c_1 - \frac{1}{c_1}\right] \\ &= 4\pi\epsilon^2 k\left[2\log\left(\frac{r-2\epsilon}{d+2\epsilon}\right) + \frac{r^2-d^2-4\epsilon(r+d)}{(d+2\epsilon)(r-2\epsilon)}\right]. \end{aligned} \tag{16}$$

The final expansion assumes that $\rho_m > c_1$, which is true for virtually all cases of interest.

Two other cases deal with discs that are partially truncated by the image boundaries. Details of these areas A_3 and A_4 are found in [14]. Because these areas contribute minimally to the overall expected area, we focus on the cases described above.

Depending on the specific values for ρ_m and c_1 we can add in the appropriate contributions of $A_1, \ldots, A_4$, together with the value for k (equation 14) to obtain an underestimate for the expected area of an error disc — the expected area of a circle in image space that will be consistent with a point expressed in terms of some affine basis. Since such discs can in general occur with equal probability anywhere in the image, the probability that a model point lies within a disc associated with an image point is simply the ratio of this area to the area of the image. Thus by normalizing these equations, we have an underestimate for the selectivity of the scheme. This leads to the following result:

Proposition 3. *Given a model basis and a fourth model point, the probability that a corresponding image basis and fourth image point will map at random to a region of α-β-space consistent with the model point and basis is given by*

$$\mu = \frac{A_1 + A_2 + A_3 + A_4}{4r^2} \tag{17}$$

where the A_i's are the areas for the four cases considered above.

This uses the upper bound on the radius of the error discs. As noted earlier, a simple lower bound can be obtained by substituting $\epsilon/2$ in place of ϵ, reflecting the use of the bound $\epsilon(1+\rho)$ in place of $2\epsilon(1+\rho)$. In this case, the bounds c_1 and c_2 will change slightly.

We can use this to compute example values for the selectivity, which depends on ρ_m (the maximum value of $|\alpha| + |\beta|$). If we allow any possible triple of points to form a basis, then ρ_m can be arbitrarily large. Consider a point p that makes an angle θ with the **u** axis, and where **u**, **v** make an angle ϕ. The value for ρ associated with the point p is

$$\frac{p}{uv|\sin\phi|}\left(u|\sin\theta| + v|\sin(\phi - \theta)|\right).$$

As ϕ approaches 0, this value becomes unbounded. We can exclude unstable bases if we set limits on the allowable range of values for ϕ, in particular, we can restrict our attention to bases with the property that

$$\phi_0 \leq \phi \leq \pi - \phi_0 \qquad \text{or} \qquad \pi + \phi_0 \leq \phi \leq 2\pi - \phi_0.$$

By applying standard minimization methods, the maximum value for ρ is given by

$$\rho_m \leq \frac{M}{m}\frac{1}{\sin\frac{\phi_0}{2}} \tag{18}$$

where m and M are the minimum and maximum distance between any two model points.

To evaluate the selectivity, we also need to know d, the length of the basis vector, $1 \leq d \leq r$. Given a specific value for d, we can compute the selectivity. To get a sense of the variation of μ, it is plotted as a function of d in figure 5, for $\epsilon = 3$.

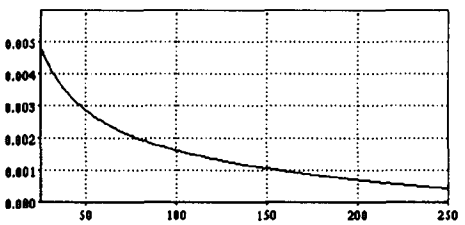

Fig. 5. Graph of selectivity μ for $\epsilon = 3$ as the basis vector length d varies.

In general, d will take on a variety of values, as the choice of basis points in the image is varied. To estimate the expected degree of selectivity, we perform the following analysis. We assume, for simplicity, that the origin of the image basis is at the center of the image. The second point used to establish the basis vector can lie anywhere in the image, with equal probability. Hence the probability distribution for d is roughly $\frac{2d}{r^2}$. We could explicitly integrate equation 17 with respect to this distribution for d to obtain an expected selectivity. This is messy, and instead we pursue two other options.

First, we can integrate this numerically for a set of examples, shown in Table 1 under the column marked *predicted*, which lists values for μ as a function of noise in the image (with an image dimension of $2r = 500$). The value of ρ_m was set using $\phi_0 = \pi/16$, and a ratio of minimum to maximum model point separation of $M/m = 10$. It should be noted that varying ϕ_0 over the range $\pi/8$ to $\pi/32$ produced results very similar to those reported in the table. As expected, the probability of a consistent match increases (selectivity decreases) with increasing error in the measurements. Thus, for ranges of parameters that one would find in many recognition situations, a considerable fraction of the space of possible α and β values are consistent with a given feature and basis.

To test the validity of our formal development, we ran a series of simulations to test the selectivity values μ predicted by equation (17). We generated sets of model and image features at random, chose bases for each at random, then checked empirically the probability that a model point, rewritten in the image basis, lay within the associated error disc of an image point. We chose to consider only cases in which the error disc fit entirely within the image bounds, since we know that our predictions are underestimates for the other cases. Table 1 summarizes the results, under the *measured* column.

Table 1. Table comparing simulated and predicted selectivities. See text for discussion.

Case	Measured	Predicted	Approximation
$\epsilon = 1$	.000116	.000117	.000118
$\epsilon = 3$	.001146	.001052	.001064
$\epsilon = 5$	.003142	.002911	.002955

Second, we can approximate the selectivity expression. By applying power series expansions and keeping only first and second order terms, we get:

$$\mu \approx \frac{k\pi\epsilon^2}{r^2}\left[\frac{17}{12} + 2\log\frac{r}{d} + \frac{r^2 - d^2}{rd}\right]. \tag{19}$$

We can find the expected value for equation 19 over the distribution for d, where $\ell \leq d \leq r$, for some minimum value ℓ. If we assume $\ell \ll r$, this expected value reduces to

$$\overline{\mu} \approx 2k\pi\left(\frac{\epsilon}{r}\right)^2\left[\frac{15}{8} - \frac{\ell}{r} - \left(\frac{\ell}{r}\right)^2 \log\frac{r}{\ell}\right]. \tag{20}$$

This predicts values close to those in Table 1, as shown in the *approximation* column.

Note that the selectivity is clearly not linear in sensor error. For a fixed size image, increasing the error ϵ by some amount should decrease the selectivity by at least a quadratic effect (perhaps more since there are higher order terms). This is reflected in Table 1. This expected value of the selectivity allows us to analyze the probability that a match will be reported at random by some recognition method that uses affine transformations. The selectivity, $\overline{\mu}$, in essence reflects the power of a given quadruple of features to distinguish a particular model. Now we consider the manner in which information from multiple quadruples is combined. This analysis differs slightly for different recognition methods. As an illustration of how the analysis applies, we consider the alignment method.

4 The Sensitivity of Alignment in the Presence of Noise

The initial version of the affine-invariant alignment method was restricted to planar objects [15], whereas later versions operate on 3D models (unlike affine hashing which uses 2D models) [16]. We consider the 2D case. The basic method is:

- Choose an ordered triple of image features and an ordered triple of model features, and hypothesize that these are in correspondence.
- Use this correspondence to compute an affine transformation mapping model into image.

- Apply this transformation to all of the remaining model features, thereby mapping them into the image.
- Search over an appropriate neighborhood about each projected model feature for a matching image feature, and count the total number of matched features.

This operation is in principle repeated for each ordered triple of model and image features, although it may be terminated after one or more matches are found, or after a certain number of triples are tried without finding a match.

We can use the expressions derived above to analyze the sensitivity of the alignment method. The key question is whether a random collection of sensor points can masquerade as a correct interpretation. In this case, we can investigate the probability of such false positive identifications as follows:

1. The selectivity of a given quadruple of points is given by $\overline{\mu}$ (equation (17)).
2. Since each model point is projected into the image, the probability that a given model point matches at least one image point is

$$p = 1 - (1 - \overline{\mu})^{s-3}$$

because the probability that a particular model point is not consistent with a particular image point is $(1 - \overline{\mu})$ and by independence, the probability that all $s - 3$ points are not consistent with this model point is $(1 - \overline{\mu})^{s-3}$.

3. The process is repeated for each model point, so the probability of exactly k of them having a match is

$$q_k = \binom{m-3}{k} p^k (1-p)^{m-3-k}. \tag{21}$$

Further, the probability of a false positive identification of size at least k is

$$w_k = 1 - \sum_{i=0}^{k-1} q_i.$$

Note that this is the probability of a false positive for a particular sensor basis and a particular model basis.

4. This process is repeated for all choices of model bases, so the probability of a false positive identification for a given sensor basis with respect to any model basis is

$$e_k = 1 - (1 - w_k)^{\binom{m}{3}}. \tag{22}$$

4.1 Testing the model

To check the correctness of our model, we ran a series of experiments based on equation 21. In particular, we used our analysis to generate a distribution for the probability of a false positive of size k, given $\epsilon = 3$ and $\phi_0 = \frac{\pi}{16}$, and using a model with 25 features and images with 25, 50, 100 and 200 features. For comparison, we also generated sets of model and image points of the same sizes, selected bases for each at random, and determined the size of vote associated with that pairing of bases. That is, for each model point (other than the basis points) we computed the affine coordinates relative to the chosen basis. Then we used the affine coordinates to determine the nominal position of an associated image point. If at least one image point was contained within a given model point's error disc, then we incremented the vote for this pairing of bases. This trial was repeated 1000 times. The results are shown in figure (6).

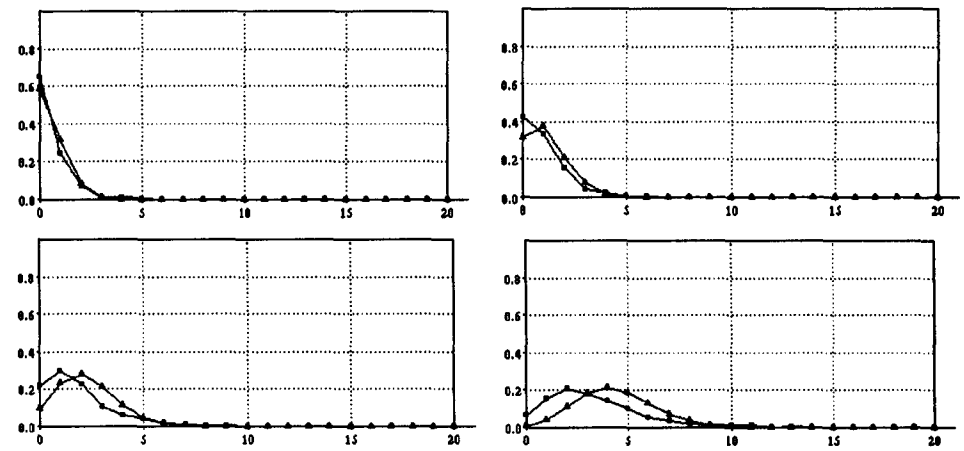

Fig. 6. Comparison of predicted and measured probabilities of false positives. Each graph compares the probability of a false peak of size k observed at random. The cases are for $m = 25$ and $s = 25$ and 50, in the top row, and $s = 100$ and 200 in the bottom row. In each case, $\epsilon = 3$. The graphs drawn with triangles show the predicted probability, while the graphs drawn with squares show the observed empirical probabilities.

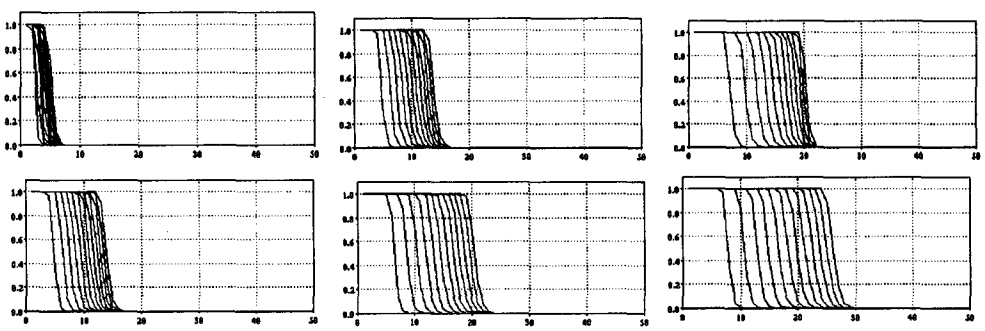

Fig. 7. Graph of probability of false positives. In each graph, vertical axis is probability of false positive of size k, horizontal axis is k. Each family of plots represents a different number of sensor features, starting with $s = 25$ for the left most plot, and increasing by increments of 25. For the three families in the top row, the model consisted of 25 features, and the sensor error was $\epsilon = 1, 3$ and $\epsilon = 5$ (from right to left). For the three families in the bottom row, ϵ was fixed at 3, and the model had 25, 38 and 50 features (from right to left).

One can see that the cases are in good agreement. In fact, our model tends to overestimate the probability of small false positives, and underestimate the probability of large false positives, so our results will tend to be conservative.

Next, what does e_k look like? As an illustration, we graph in figure (7) the probability of a false positive based on equation (22). In particular, we use a selectivity based on $\epsilon = 3$, obtained from Table 1, and plot the value of e_k for an object with 25, 38 or 50 model features, for different values of k and a given number of sensor features s. This is graphed in figure (7). The process was repeated for different numbers of sensor features

s, generating the family of graphs in the figure.

In figure (7), we show the false positive rate as the error rate changes. Each figure plots the false positive rate, for model features $m = 25$, and for sensor features varying from $s = 25$ to $s = 200$ by increments of 25. The individual plots are for varying numbers of sensory features, and the process is repeated for changes in the bound on the sensor error, given a fixed threshold on angle of $\phi_0 = \pi/16$. If the error is very small, the method performs well, i.e. the probability of a false positive rapidly drops to zero even for small numbers of model features. As the error increases, however, the probability of a false positive rapidly increases, as can be seen by comparing different families of plots in figure (7). Note that the best possible correct solution would be for $k = 22$.

While we have used our methods to analyse the alignment method, very similar results hold for the case of geometric hashing. For details, see [14].

5 Summary

The computation of an affine-invariant representation in terms of a coordinate frame $(\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3)$ has been used in a number of model-based recognition methods. Nearly all of these recognition methods were developed assuming no uncertainty in the sensory data, and then various heuristics were used to allow for error in the locations of sensed points. In this paper we have formally examined the effect of sensory uncertainty on such recognition methods. This analysis involves considering both the Euclidean plane used by the alignment method, and the space of affine-invariant (α, β) coordinates used by the geometric hashing method. Our analysis models each sensor point in terms a *disc* of possible locations, where the size of this disc is bounded by an uncertainty factor, ϵ.

Under the bounded uncertainty error model, in the Euclidean space the set of possible values for a given point $\mathbf{x}$ and a basis $(\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3)$ forms a disc whose radius is bounded by $r = k\epsilon(1 + |\alpha| + |\beta|)$, where $1 \leq k \leq 2$. That is, assuming that each image point has a sensing uncertainty of magnitude ϵ, the range of image locations that are consistent with $\mathbf{x}$ forms a circular region. In the α-β-space, the set of possible values of the affine coordinates of a point $\mathbf{x}$ in terms of a basis $(\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3)$ forms an ellipse (except in degenerate cases). The area, center and orientation of this ellipse are given by somewhat complicated expressions that depend on the actual configuration of the basis points.

The most important consequence of our analysis is that the set of possible values in the α-β-plane *cannot* be computed independent of the actual locations of the model or image basis points. This means that the table constructed by the geometric hashing method can only approximate the correct values, because the locations of the image points are not known at construction time. We further find for even moderately large positional uncertainty, methods that use affine transformations have a substantial probability of false positive matches. These methods only check the consistency of each matched point with a set of basis matches. They do not ensure the global consistency of all matched points. Our results suggest that such methods will require that a substantial number of hypothesized matches be ruled out by some subsequent verification stage.

References

1. Ballard, D.H., 1981, "Generalizing the Hough Transform to Detect Arbitrary Patterns," *Pattern Recognition* **13**(2): 111–122.
2. Basri, R. & S. Ullman, 1988, "The Alignment of Objects with Smooth Surfaces," *Second Int. Conf. Comp. Vision*, 482–488.

3. Besl, P.J. & R.C. Jain, 1985, "Three-dimensional Object Recognition," *ACM Computing Surveys*, **17**(1):75–154.
4. Costa, M., R.M. Haralick & L.G. Shapiro, 1990, "Optimal Affine-Invariant Point Matching," *Proc. 6th Israel Conf. on AI*, pp. 35–61.
5. Cyganski, D. & J.A. Orr, 1985, "Applications of Tensor Theory to Object Recognition and Orientation Determination", *IEEE Trans. PAMI*, **7**(6):662–673.
6. Efimov, N.V., 1980, *Higher Geometry*, translated by P.C. Sinha. Mir Publishers, Moscow.
7. Ellis, R.E., 1989, "Uncertainty Estimates for Polyhedral Object Recognition," *IEEE Int. Conf. Rob. Aut.*, pp. 348–353.
8. Forsythe, D., J.L. Mundy, A. Zisserman, C. Coelho, A. Heller, & C. Rothwell, 1991, "Invariant Descriptors for 3-D Object Recognition and Pose", *IEEE Trans. PAMI*, **13**(10):971–991.
9. Grimson, W.E.L., 1990, *Object Recognition by Computer: The role of geometric constraints*, MIT Press, Cambridge.
10. Grimson, W.E.L., 1990, "The Combinatorics of Heuristic Search Termination for Object Recognition in Cluttered Environments," *First Europ. Conf. on Comp. Vis.*, pp. 552–556.
11. Grimson, W.E.L. & D.P. Huttenlocher, 1990, "On the Sensitivity of the Hough Transform for Object Recognition," *IEEE Trans. PAMI* **12**(3):255–274.
12. Grimson, W.E.L. & D.P. Huttenlocher, 1991, "On the Verification of Hypothesized Matches in Model-Based Recognition", *IEEE Trans. PAMI* **13**(12):1201–1213.
13. Grimson, W.E.L. & D.P. Huttenlocher, 1990, "On the Sensitivity of Geometric Hashing", *Proc. Third Int. Conf. Comp. Vision*, pp. 334–338.
14. Grimson, W.E.L., D.P. Huttenlocher, & D.W. Jacobs, 1991, "Affine Matching With Bounded Sensor Error: A Study of Geometric Hashing & Alignment," MIT AI Lab Memo 1250.
15. Huttenlocher, D.P. and S. Ullman, 1987, "Object Recognition Using Alignment", *Proc. First Int. Conf. Comp. Vision*, pp. 102-111.
16. Huttenlocher, D.P. & S. Ullman, 1990, "Recognizing Solid Objects by Alignment with an Image," *Inter. Journ. Comp. Vision* **5**(2):195–212.
17. Jacobs, D., 1991, "Optimal Matching of Planar Models in 3D Scenes," *IEEE Conf. Comp. Vis. and Patt. Recog.* pp. 269–274.
18. Klein, F., 1939, *Elementary Mathematics from an Advanced Standpoint: Geometry*, MacMillan, New York.
19. Korn, G.A. & T.M. Korn, 1968, *Mathematical Handbook for Scientists and Engineers*, McGraw-Hill, New York.
20. Lamdan, Y., J.T. Schwartz & H.J. Wolfson, 1988, "Object Recognition by Affine Invariant Matching," *IEEE Conf. Comp. Vis. and Patt. Recog.* pp. 335–344.
21. Lamdan, Y., J.T. Schwartz & H.J. Wolfson, 1990, "Affine Invariant Model-Based Object Recognition," *IEEE Trans. Rob. Aut.*, vol. 6, pp. 578–589.
22. Lamdan, Y. & H.J. Wolfson, 1988, "Geometric Hashing: A General and Efficient Model-Based Recognition Scheme," *Second Int. Conf. Comp. Vis.* pp. 238–249.
23. Lamdan, Y. & H.J. Wolfson, 1991, "On the Error Analysis of 'Geometric Hashing'," *IEEE Conf. Comp. Vis. and Patt. Recog.* pp. 22–27.
24. Thompson, D. & J.L. Mundy, 1987, "Three-Dimensional Model Matching From an Unconstrained Viewpoint", *Proc. IEEE Conf. Rob. Aut.* pp. 280.
25. Van Gool, L., P. Kempenaers & A. Oosterlinck, 1991, "Recognition and Semi-Differential Invariants," *IEEE Conf. Comp. Vis. and Patt. Recog.* pp. 454–460.
26. Wayner, P.C., 1991, "Efficiently Using Invariant Theory for Model-based Matching," *IEEE Conf. Comp. Vis. and Patt. Recog.* pp. 473–478.
27. Weiss, I., 1988, "Projective Invariants of Shape," *DARPA IU Workshop* pp. 1125–1134.
28. Wolfson, H.J., 1990, "Model Based Object Recognition by Geometric Hashing," *First Europ. Conf. Comp. Vis.* pp. 526–536.

This article was processed using the LaTeX macro package with ECCV92 style