MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo No. 1476 December, 1994

Uncertainty Propagation in Model-Based Recognition

D. W. Jacobs **T. D. Alter**

This publication can be retrieved by anonymous ftp to publications.ai.mit.edu.

Abstract

Building robust recognition systems requires a careful understanding of the effects of error in sensed features. In model-based recognition, matches between model features and sensed image features typically are used to compute a model pose and then project the unmatched model features into the image. The error in the image features results in uncertainty in the projected model features. We first show how error propagates when poses are based on three pairs of model and image points. In particular, we show how to simply and efficiently compute the region in the image where an unmatched model point might appear, for both Gaussian and bounded error in the detection of image points, and for both scaled-orthographic and perspective projection models. This result applies to objects that are fully three-dimensional, where past results considered only two-dimensional objects. The result is based on an approximation that accurately linearizes the relationship between matched image points and unmatched, projected model points. Secondly, based on the linear approximation, we show how we can utilize *linear programming* to compute the propagated error region for any number of initial matches. Finally, we use these results to extend, from two-dimensional to three-dimensional objects, robust implementations of *alignment*, *interpretation-tree search*, and *transformation clustering*.

This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology and at NEC Research Institute. DWJ was supported by NEC Research Institute. TDA was supported by Office of Naval Research AASERT Reference No. N00014-94-1-0128 and by NEC Research Insitute. Support for the Artificial Intelligence Laboratory's research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-91-J-4038.

1 Introduction

Given a correspondence between a set of image features and model features, a general problem in recognition is to evaluate the correspondence and improve it if necessary. For instance, for object recognition the model may be a sparse set of 3D points and line segments. For aerial images, the model may be a terrain elevation map that includes the world locations of a small set of landmarks. In some applications, a user may supply the initial correspondence, leaving the computer to estimate and refine the model pose (position and orientation). In other cases, the computer must find the initial correspondence as well; this may be done through a combination of grouping, indexing, and raw search. Important computations involved in evaluating and improving the correspondence include (1) deciding whether the correspondence provides an accurate alignment, (2) determining which image features could correspond to each unmatched model feature, and (3) choosing a new match to extend the correspondence. These computations are intertwined with the issue of error propagation, that is, the issue of how error in a set of matched image features propagates to uncertainty in the predicted image locations of the remaining model features. We call these predicted image locations the *uncertainty regions* of the model features, and we derive either bounds on these regions or probability distributions on them, depending on our model of error.

There are several reasons why it is useful to carefully understand the propagation of uncertainty, as opposed to assuming some small, simple uncertainty region and using it in all cases. First, as we will show, uncertainty regions can vary quite a bit in size, and may be quite large for the predicted model features, resulting in many candidate image features for each prediction. In particular, grouping techniques commonly find image features that are close together on an object (e.g., [11, 8, 25, 31, 27, 29]), and we will see that this easily can lead to large uncertainty regions. Even when the matched features are far apart in the image, the uncertainty regions of the unmatched points may still be large, due to the depth of the 3D model. Second, both when the image features are nearby and when they are far apart, there are situations in which the pose of the model is unstable, and the uncertainty regions assume surprising shapes. By understanding the propagation of uncertainty, then, we can determine exactly where to look for features, and we can evaluate the stability of the pose produced by the initial correspondence.

1.1 Summary of Results

Given a set of matched image and model points, we determine an unmatched model point's uncertainty region. We consider this problem for the case in which correspondences are based on point features. We handle both scaled-orthographic and perspective projection models. We also consider two different models of error. First, we consider image points detected with errors that have known, independent Gaussian distributions. Second, we consider a bounded error model, in which we suppose that the error distributions are unknown. In this case we make only the weak assumption that the magnitude of the error vectors can be bounded by some maximum number of pixels ϵ. Given no other information, Gaussians may be the preferred error distribution, since image features are displaced by a sum of error vectors, incurred over a series of processes such as digitization, smoothing, and edge detection. A bounded error model may be useful, however, when errors contain a consistent bias that results in distributions that are significantly skewed from Gaussian. In the first case, we show how Gaussian error in matched image points propagates to an uncertainty region with a Gaussian distribution for an unmatched point. In the second case, we show how bounded error in image points propagates to a bounded uncertainty region describing the possible location of an additional model point.

First we compute the uncertainty regions for sets of three matched points. We derive a simple linear expression that approximates the relationship between the matched and unmatched points. This relationship allows us to show that, for bounded error, the uncertainty region for a fourth point is circular, and to derive analytic expressions for the center and radius of the circle. For Gaussian error, this relationship implies that the propagated distribution of uncertainty is also Gaussian, and provides analytic expressions for the center and standard deviation. We perform experiments to verify that these expressions are accurate for the amount of error that is of interest in most recognition applications.

We also take advantage of the linear relationship by introducing a new algorithm that allows us to determine the uncertainty region for any number of matched points. To do this we approximate our bounded error regions with convex polygons, and then show that we can use linear programming to derive a convex polygon that describes the uncertainty region of the unmatched model point. We experiment with both synthetic images and a real image to observe the accuracy of the uncertainty regions that we compute, and to determine the extent to which they shrink as we match more points.

Finally, we show how to extend previous work for linear projection models to the cases of scaled-orthographic and perspective projections. Using the linear approximation we show that we can use Baird's [6] algorithm to tell whether a set of matches between image and model points are geometrically consistent, and that we can apply Cass' [12] and Breuel's [10] algorithms to find, in polynomial time, the model pose that aligns the maximum number of model and image features to within error bounds. We also extend Jacobs' [28] and Sarachik and Grimson's [39] planar alignment algorithms to 3D objects.

1.2 Projection Models

For reference, we review the models of projection that we refer to in this paper. For perspective projection, we can write the corresponding image position (x, y) of a 3D model point $(\overline{x}, \overline{y}, \overline{z})$ in terms of a 3D, rigid rotation matrix $\mathbf{R}$, a 3D translation vector $\vec{u}$, and a camera focal

length f. Letting r_{ij} be the elements of $\mathbf{R}$, we have

$$x = f\frac{r_{11}\overline{x} + r_{12}\overline{y} + r_{13}\overline{z} + u_x}{r_{31}\overline{x} + r_{32}\overline{y} + r_{33}\overline{z} + u_z}, \quad (1)$$

$$y = f\frac{r_{21}\overline{x} + r_{22}\overline{y} + r_{23}\overline{z} + u_y}{r_{31}\overline{x} + r_{32}\overline{y} + r_{33}\overline{z} + u_z}, \quad (2)$$

where the rows of $\mathbf{R}$ are orthonormal, and where we assume the origin is at the center of projection. When the focal length f is known, there are six degrees of freedom, and consequently three corresponding model and image points are "minimal" to determine the transformation. Given three corresponding points, there exist up to four solutions for the model pose [17].

This paper extensively considers scaled-orthographic (also known as weak-perspective) projection, in which a 3D object is scaled down and projected orthographically into the image. This projection model is appropriate when the camera is far from the objects being viewed with respect to their sizes. In this case, the image position of $(\overline{x}, \overline{y}, \overline{z})$ can be written in terms of the first two rows of a scaled, 3D rotation matrix, $\mathbf{S} = s\mathbf{R}$, and of a scaled, 3D translation vector, $\vec{b}$. Letting s_{ij} be the elements of $\mathbf{S}$, we have

$$x = s_{11}\overline{x} + s_{12}\overline{y} + s_{13}\overline{z} + b_x, \quad (3)$$

$$y = s_{21}\overline{x} + s_{22}\overline{y} + s_{23}\overline{z} + b_y, \quad (4)$$

where $\| \ (s_{11}, s_{12}, s_{13}) \ \| = \| \ (s_{21}, s_{22}, s_{23}) \ \|$ and $(s_{11}, s_{12}, s_{13}) \cdot (s_{21}, s_{22}, s_{23}) = 0$. There are six degrees of freedom in the scaled-orthographic model-to-image transformation, and consequently three corresponding points are minimal to determine the transformation. Given three corresponding points, the transformation always exists if the model points are not collinear and it generally has two solutions [27, 2]; in particular, the scale factor and translation are always unique, and the rigid rotation matrix is unique up to a reflection of the rotated model about a plane parallel to the image.

For 3D linear projection, we remove the two non-linear constraints on the rotation parameters in the scaled-orthographic projection model. This transformation is equivalent to applying a scaled-orthographic transformation to the model, and then applying a scaled-orthographic transformation to the resulting image; in total, this is like taking a picture of a photograph [29]. There are eight degrees of freedom in linear projection, and four corresponding points are minimal to determine the transformation. Given a minimal set of matches, this is the only transformation of the three in which the unmatched model points can be written *linearly* in terms of the matched image points. In particular, let the four image and model points be (x_i, y_i) and $(\overline{x}_i, \overline{y}_i, \overline{z}_i)$, respectively, for $i = 1, 2, 3, 4$. Then we can obtain the first row of the transformation by solving

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} \overline{x}_1 & \overline{y}_1 & \overline{z}_1 & 1 \\ \overline{x}_2 & \overline{y}_2 & \overline{z}_2 & 1 \\ \overline{x}_3 & \overline{y}_3 & \overline{z}_3 & 1 \\ \overline{x}_4 & \overline{y}_4 & \overline{z}_4 & 1 \end{bmatrix} \begin{bmatrix} s_{11} \\ s_{12} \\ s_{13} \\ b_x \end{bmatrix}. \quad (5)$$

A similar equation holds for the second row of the transformation. These equations give linear expressions for the transformation parameters in terms of the image point coordinates. Since multiplying a matrix by a vector is a linear operation, applying the computed transformation to any unmatched model point gives a linear expression for the model point's image position in terms of the matched image points.

1.3 Background

Due to the value of top-down knowledge in model-based vision, it is common to generate hypotheses about an object's pose based on a small amount of information, and then to look for evidence to confirm or reject the hypotheses. In the *alignment* approach, a small number of image features are matched to model features to determine the object's pose. This pose is used to project additional model features into the image, which are matched to nearby image features for verification (e.g., Roberts [37], Clark et al. [13], Fischler and Bolles [17], Lowe [34], Ayache and Faugeras [5], Huttenlocher and Ullman [27]). In *interpretation-tree search*, additional matches between model and image features are then used to look for more matches, backtracking if enough valid matches cannot be found (e.g., Bolles and Cain [8], Goad [18], Grimson and Lozano-Pérez [23], Horaud [25]). To obtain the object's pose, some approaches use minimal sets of matches between model and image features (e.g., Clark et al. [13], Fischler and Bolles [17], Ayache and Faugeras [5], Horaud [25], Huttenlocher and Ullman [27]). Other approaches use indexing to match more than the minimal number before looking for confirming features (e.g., Rothwell et al. [38], Thompson and Mundy [43], Lamdan et al. [32], Jacobs [29]).

Most recognition systems take an ad-hoc approach to the problem of accounting for the effects of sensing error on the projected positions of unmatched model features. Some systems match projected model features to image features if they are separated by a distance that is less than some threshold (e.g., Clark et al. [13], Fischler and Bolles [17], Brooks [11], Bolles and Cain [8], Huttenlocher and Ullman [27]). Other systems rank the unmatched image features using heuristics involving distance and orientation, and then pick the feature with highest rank (e.g., Ayache and Faugeras [5], Lowe [34]). Many questions remain concerning the performance of these systems. For example, although we know the minimal number of features needed to generate a model pose, we do not know how accurate the pose must be to allow us to identify the object. In addition, some authors stress the importance of using a minimal set of features [17, 27], while others contend that this will not produce a sufficiently accurate pose for recognition [34]. It is in general difficult to characterize the conditions under which these systems will succeed or fail, or to evaluate the relative effectiveness of the different strategies for recognition, or to understand the extent to which each approach makes the best possible use of the information available. A careful understanding of the effects of sensing error is a prerequisite to doing all of these.

1.3.1 Two-dimensional objects

Recently, there has been considerable effort aimed at better understanding the effects of error on the match-

ing process. Some of this work attempts to design algorithms that are guaranteed to perform well in the presence of error (e.g., Baird [6], Cass [12], Breuel [10]), but most relevant to this paper is work that also examines the propagation of error in recognition systems.

Huttenlocher [26] examined the effects of bounded error on the alignment approach to recognition. This analysis considered planar objects viewed from arbitrary 3D positions, assuming scaled-orthographic projection. Pose was determined by matching three model and image points. For some situations, Huttenlocher placed approximate bounds on the uncertainty regions.

Subsequently, Jacobs [28] showed that the true uncertainty regions are discs, and gave analytic expressions for their centers and radii. These regions are circular because in this case the projection model is linear in such a way that error in any of the three matched image points causes error in a projected model point that is identical but scaled by a constant factor. This constant factor depends on the model structure, but not on the viewpoint. Consequently, the sizes of the uncertainty regions are independent of how far apart in the image are the three matched points, which means the uncertainty is independent of the pose of the model. Jacobs' result was used by Grimson et al. [22] to analyze the false-positive sensitivity of planar alignment.

A number of researchers have also considered the effect of Gaussian error on alignment methods. As mentioned above, for planar objects, each predicted model point can be written as a linear combination of the matched image points. Therefore, Gaussian error in the image points leads to Gaussian uncertainty in every predicted point (e.g., [42]). Sarachik and Grimson [39] used this observation to propose a new method of performing and evaluating alignment approaches to recognition. Beveridge et al. [7] use a robust method to evaluate particular model poses.

Error propagation has also been studied in the context of Geometric Hashing approaches to recognition. Costa et al. [15] considered the distribution of uncertainty regions in terms of the affine invariant parameters that describe the image points. Rigoutsos and Hummel [35, 36] also considered this issue for Gaussian and uniform error. Both Costa et al. and Rigoutsos and Hummel then considered the implications of these results for recognition schemes. Lamdan and Wolfson [33] considered the related problem of determining when three image points provide an unstable basis for Geometric Hashing. Grimson and Huttenlocher [20] considered the effects of bounded error on Geometric Hashing, and provided loose bounds on this effect. Jacobs [28] determined exactly how bounded error effects Geometric Hashing indices. Grimson et al. [22] then further developed this result and used it to analyze the performance of Geometric Hashing algorithms. Sarachik and Grimson's [39] results also apply to Geometric Hashing.

1.3.2 Three-dimensional objects

Error propagation is more complex in recognition systems that deal with fully three-dimensional objects. Bolles et al. [9] studied how error propagates from the parameters of a model-to-image transformation to the predicted model points. Bolles et al. assumed that the errors in the parameters were independent and normally-distributed and that estimates of the distributions would be available. Unlike other previous work, Bolles et al. dealt with perspective projection, which made the relationship between the error vectors in the transformation parameters and the predicted points non-linear. In fact, their analysis is the most similar to our own, because they took a (first-order) approximation that linearizes the error-vector relationship. As a result they obtained Gaussian uncertainty distributions. The main difference with our work, in addition to our treatment of bounded error, is that we will let the error be in the matched image points, instead of assuming we know the distributions for all of the transformation parameters. Furthermore, we will derive direct expressions for the predicted points in terms of the matched points, so that we do not explicitly go through a rigid transformation.

Recently, Grimson et al. [21] presented a formal analysis of error propagation starting from the matched image points, for three-dimensional objects. They considered scaled-orthographic projection and bounded, circular error. Starting from three matched points, they provided a numerical method of bounding the uncertainty in the transformation parameters. Then they used the bounds on the parameters to obtain complicated, loose bounds on the uncertainty regions of the predicted points. Via these bounds, they analyzed the false-positive sensitivity of 3D-from-2D alignment and transformation clustering, in the domain of point features. The numerical technique is less practical, however, for use at run-time in a recognition system.

Using the same projection and error models as Grimson et al. [21], Alter and Grimson [4] presented experiments that show that the true uncertainty regions tend to be circular to a good approximation, and presented a numeric method for more accurately bounding the uncertainty regions. This technique was used to study again the false-positive sensitivity of 3D-from-2D alignment, except also using line features for verification. Alter and Grimson demonstrated that using points for generating hypotheses and lines for verification could lead to robust recognition. As before, the numerical error-propagation technique is less practical for a real-time system. Furthermore, the two weak-perspective solutions lead to two distinct uncertainty regions, which is not true when the model is planar. Alter and Grimson's technique sometimes performed poorly when the two regions overlapped, because it had difficulty distinguishing them.

Also for 3D objects, Weinshall and Basri [46] provided analytic bounds on the amount of error in a least-squares solution that is used to match four model and image points. This is useful because, currently, the least-squares solution itself can be found only through iterative methods.

For both 3D and 2D objects, Wells [47, 48] used a Bayesian approach and Gaussian error assumptions to derive an evaluation function that measures the likelihood of any given pose. Wells then used heuristic search

and gradient descent methods to find the most probable pose.

Finally, there has been a great deal of work on finding a pose that minimizes error, when enough image and model features have been matched to overdetermine the pose. Some of this work analyzes the effect that errors in image features have on the accuracy of the resulting pose, including Kumar and Hanson [30] and Hel-Or and Werman [24]. The work of Hel-Or and Werman is particularly relevant to us, because they also consider how error propagates through the pose to the projections of unmatched feature points. Assuming Gaussian error, they use an extended Kalman filter to find the minimal error pose resulting from a match between any number of image and model points. The Kalman filter then allows them to compute a Mahalanobis distance that indicates the likelihood that error can account for the apparent deviation between a projected model point and a potentially matching image point.

In summary, there are simple analytic solutions for how error propagates from three matched image points, when the objects are two-dimensional and undergo scaled-orthographic projection. This is true both when the image-point error is bounded by circles and when it is normally distributed. In the case of circular error, every propagated uncertainty region is a circle, whose size is independent of the camera viewpoint.

For three-dimensional objects, it appears empirically that circular error again propagates to circular uncertainty regions. Nevertheless, there is no analytic solution, which would be preferred for building an efficient system. As well, current numerical solutions either significantly overestimate the uncertainty regions or can break down when the two regions that arise from the two weak-perspective solutions overlap. Further, it is not known whether the uncertainty regions are exactly or approximately circles, or whether the sizes of the regions depend on the viewpoint. If the regions are circles only approximately, one would like to know which configurations of the model and image points cause the regions to deviate from circularity. Although much progress has been made in understanding the effects of propagated error, there are significant problems that are not yet understood.

Finally, there have been a number of sensitivity analyses that determine the susceptibility of recognition systems to false-positive errors. Most of these analyses are restricted to two-dimensional objects, because this is where error propagation is most readily understood. Nonetheless, there do exist sensitivity analyses for three-dimensional objects, which use numerical techniques to get a handle on the propagated error.

2 Fourth-Point Uncertainty Region

In this section, we address the following problem: Given exactly three matching point pairs, $(\vec{i}_0, \vec{m}_0)$, $(\vec{i}_1, \vec{m}_1)$, and $(\vec{i}_2, \vec{m}_2)$, where the locations of $\vec{i}_0$, $\vec{i}_1$, and $\vec{i}_2$ contain small amounts of error, what is the error in the computed image position of a fourth model point, $\vec{m}_3$? This section presents an analytic solution to this problem, which