MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo No. 1044 May, 1988

On the Sensitivity of the Hough Transform for Object Recognition

W. Eric L. Grimson
Daniel P. Huttenlocher

Abstract. Object recognition from sensory data involves, in part, determining the pose of a model with respect to a scene. A common method for finding an object's pose is the generalized Hough transform, which accumulates evidence for possible coordinate transformations in a parameter space whose axes are the quantized transformation parameters. Large clusters of similar transformations in that space are taken as evidence of a correct match. In this article, we provide a theoretical analysis of the behavior of such methods. We derive bounds on the set of transformations consistent with each pairing of data and model features, in the presence of noise and occlusion in the image. We also provide bounds on the likelihood of false peaks in the parameter space, as a function of noise, occlusion, and tessellation effects. We argue that blithely applying such methods to complex recognition tasks is a risky proposition, as the probability of false positives can be very high.

Acknowledgements: This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's artificial intelligence research is provided in part by an Office of Naval Research University Research Initiative grant under contract N00014-86-K-0685, and in part by the Advanced Research Projects Agency of the Department of Defense under Army contract number DACA76-85-C-0010 and under Office of Naval Research contract N00014-85-K-0124. WELG is supported in part by the Matsushita Chair of Electrical Engineering.

1. Introduction

The recognition of partially occluded objects from noisy data is an important component of many problems in vision and robotics. Recognizing an object generally entails finding a matching between elements of an object model and instances of those elements in the data, and thereby recovering a transformation that maps a model of the object onto a portion of an image. There are a variety of approaches to the problem of finding possible transformations (see [Besl and Jain 85], [Chin and Dyer 86] for recent surveys), a common subclass of which are based on transformation clustering. The generalized Hough transform [Ballard 81] [Davis 82], or related parameter hashing techniques, are often used to perform the transformation clustering (e.g., [Thompson and Mundy 87] [Silberberg et al. 84] [Silberberg et al. 86] [Lamdan et al. 87] [Turney et al. 85]).

In this paper, we consider the robustness of clustering methods based on variations of the generalized Hough transform. We investigate the power of such methods to distinguish clusters that are due to a correct matching of image and model features from those that occur at random. We find that the methods work well as long as the correct match accounts for both much of the model and much of the sensory data. For moderate levels of sensor noise, occlusion, and image clutter, however, the methods can hypothesize many false solutions, and their effectiveness is dramatically reduced.

The idea underlying transformation clustering methods is to accumulate independent pieces of evidence for a match. Each pair of model and image features (such as edges or vertices) defines a range of possible transformations from a model to an image. In the case of rigid objects, each transformation consists of a translation and rotation from the model coordinate system to the image coordinate system, and thus specifies the pose of the model with respect to the image. The uncertainty in the range of possible transformations depends on the type of feature, and on the degree of accuracy in the measurement of the features.

Ranges of transformations consistent with a feature pair are computed for all pairs of model and image features. Those pairs that are part of the same correct match of a model to an image will result in approximately the same transformations. Random pairs of model and image features, on the other hand, will result in randomly distributed transformations. Thus a cluster of similar transformations is assumed to correspond to a correct match. The validity of this assumption, however, depends on there being a low likelihood that random clusters will be as large as those clusters resulting from correct matches.

Two techniques are commonly used to find clusters in an n-dimensional parameter space: k-means clustering and the generalized Hough transform. These techniques both start with a set, P, of parameter vectors, or points in the n-dimensional parameter space, and yield a set of subsets of P, where each subset is a cluster of similar parameter vectors. In transformation clustering approaches to recogni-

tion, each dimension of the parameter space, P, corresponds to a component of the transformation from a model to an image.

The k-means method is an iterative technique that starts by dividing the parameter vectors into k groups, and then iteratively moves vectors from one group to another in order to minimize the total distance between elements in each group. The k-means clustering algorithm requires a distance metric to be defined for comparing any two parameter vectors. In the case of transformations from a model to an image it is difficult to define an appropriate distance metric, because the parameter space consists of both translations and rotations, which are not directly comparable. A further limitation of the approach is that some pre-defined number of clusters, k, must be used. Thus the system must have a reasonable guess of how many meaningful clusters there are (i.e. how many object instances are in an image).

Rather than using the k-means method, object recognition systems tend to cluster transformations using the generalized Hough transform. The Hough technique works by quantizing the parameter space into discrete n-dimensional buckets. Each parameter vector is entered into a bucket by quantizing its n parameter values and using them as indices into an n-dimensional table. The quantization will generally map similar parameter vectors into the same bucket. Hence, the search for large clusters of similar transformations simply requires examining each bucket to find those buckets with the most entries.

The remainder of this paper considers the effectiveness of using the generalized Hough transform to find clusters of similar transformations in order to match a model to an image. Three central questions are addressed in this investigation:

1. What is the range of transformations specified by a given pairing of model and image features?
2. How many Hough buckets are specified by such a range of transformations?
3. How many model-image pairings are likely to fall into the same Hough bucket at random?

The first two questions are considered in Section 3, which analyzes the amount of uncertainty involved in computing a two-dimensional transformation from a model to an image, using either pairs of straight edge fragments or pairs of vertices. In Section 4, the generalized Hough transform is modeled as an occupancy problem, in order to estimate the size clusters that are likely to occur at random. This analysis makes use of the analytic results from Section 3, as well as some empirical data from existing recognition systems. We find that for a wide variety of tasks, clusters occurring at random are as large in size as those that are due to a correct match. Thus for these tasks, the generalized Hough method is not a good technique for finding correct matches of a model to an image.

A number of other authors have considered aspects of the noise sensitivity of the Hough transform, usually in the case of detecting lines or other simple curves in noisy images [Shapiro 75, Maitre 76, Cohen and Toussaint 77, Shapiro and Iannino

79, Alagar and Thiel 81, van Veen and Groen 81]. Brown [1983] has considered the noise properties of more general applications of the Hough transform, by treating the problem as one of signal processing. In this article, we take a different approach, using discrete combinatorial tools to analyze the problem.

Before addressing the three questions posed above, the next section considers the generalized Hough transform in more detail. Some of the limitations of the simplest formulation of the technique are considered, along with the methods that are generally used to overcome those limitations. Unfortunately, these methods turn out to the increase the likelihood that many model-data pairings will fall into the same Hough bucket at random.

2. Parameter hashing: the generalized Hough transform

The generalized Hough transform finds possible solutions to the object pose problem by searching for large clusters of evidence in a discrete version of a parameter space. A parameter vector, $\mathbf{p}$, represents a point in an n-dimensional space, $\mathcal{P}$. Each point in $\mathcal{P}$ maps to a point in the n-dimensional discrete Hough space, $\mathcal{H}$, that is specified by quantizing each of the n components of $\mathbf{p}$. The Hough transform method is often also referred to as parameter hashing, because each quantized parameter value is a hash key. Implementations of the Hough method generally use an n-dimensional table to represent $\mathcal{H}$, and refer to the entries in the table as buckets.

When the generalized Hough method is used for transformation clustering, each dimension of the parameter space, $\mathcal{P}$, corresponds to a component of the transformation from a model to an image. If the coordinate system of the image measurements is denoted by $\mathcal{I}$, and the model coordinate system is denoted by $\mathcal{M}$, then $\mathcal{P}$ is the space of mappings from $\mathcal{M}$ to $\mathcal{I}$.

For each pair of model and image features, the range of possible transformations is computed. This set of transformations defines a region, $T \subseteq \mathcal{P}$. The quantized values of this n-dimensional volume, T, are used to enter the model-image pairing into all the buckets in $\mathcal{H}$ that intersect the range of possible transformations. Those model-image pairings that fall into the same quantization bucket define a cluster of similar transformations. It is assumed that the large clusters will identify correct transformations from a model to an image. Thus recognition consists of searching the n-dimensional discrete table (the space $\mathcal{H}$) for those buckets with a large number of entries.

As an example, suppose that a model consists of linear segments, and the sensory data has been processed to produce comparable linear segments. Suppose there are m different model fragments, and s sensory fragments. Each sensory measurement taken from $\mathcal{I}$ is matched in turn with each model fragment, for a total of ms model-data pairings. Consider the pairing of data edge j with model edge

J. We can compute the transformation required to bring the model fragment into correspondence with the data fragment. In two dimensions, this transformation can be defined as the angle of rotation θ_{jJ} needed to align the tangents of the two fragments, and the two dimensional translation $\mathbf{t}_{jJ}$ needed to then align the rotated model edge with the data edge.

In the case of no uncertainty, $(\theta_{jJ}, \mathbf{t}_{jJ})$ exactly defines the transformation associated with the data-model pair, jJ. This transformation $(\theta_{jJ}, \mathbf{t}_{jJ})$ is represented by a point in the three-dimensional transform space $\mathcal{P}$. If there is sensor error or partial occlusion, then the pairing jJ defines a range of a possible transformations, represented by a volume in $\mathcal{P}$. The corresponding parameters θ_{jJ} and $\mathbf{t}_{jJ}$ are quantized, and used to enter the pairing jJ into those buckets of the three-dimensional Hough table that intersect the volume in $\mathcal{P}$.

There are three problems with the generalized Hough method as presented:

1. Similar parameter vectors will end up in different buckets if they are on different sides of a quantization boundary. This problem is exacerbated by uncertainty in the parameter values.
2. For high dimensionality parameter spaces, the table can get very large, making the search for large clusters cumbersome.
3. The likelihood of large clusters occurring at random can be quite high, because the quantization integrates noise by collecting together all the random events within a bucket. The likelihood depends on the ratio of the number of parameter vectors to the number of buckets.

Two methods are often used to ensure that similar parameter vectors end up in the same cluster. The first method computes clusters over a local k^n neighborhood of buckets in the Hough table, rather than a single bucket. Generally a 3^n neighborhood is used, so that any transformations that are within one bucket of each other, along any dimension, will be clustered together. The second method computes the range of possible buckets that each data-model pairing could fall in, and enters it into each of these buckets. Both methods have the effect of increasing the number of parameter vectors entered into the table, thereby increasing the likelihood that large clusters will occur at random.

Reducing the size of the table, so that search space is of a tractable size, increases the likelihood of large clusters occurring at random. The fewer buckets there are, the more likely that many parameter vectors will fall into the same bucket at random. Most systems that use the generalized Hough technique for clustering in high dimensional parameter spaces (such as six degree of freedom three-dimensional recognition tasks) use only a subset of the parameters to define the Hough table. This greatly reduces the size of the table, but at the same time greatly increases the chance of large random clusters.

Thus the techniques used to address the first two problems exacerbate the third problem. It is this problem that we analyze using a combinatoric model in Section

4. First, however, we derive bounds on the number of Hough buckets specified by each pairing of model and image data elements.

3. Two dimensional noise analysis

This section addresses the issue of what we will term the *redundancy factor* of entering transformations into the table. That is, how many different buckets in the Hough table can the same model-data pairing specify? This depends on the dimensionality of the sensory data, the dimensionality of the transformation from a model to an image, the coarseness of the tessellation of the Hough space, and the expected amount of noise in the sensory measurements.

To determine the redundancy factor we need a method for estimating the set of transformations consistent with a data-model pairing, under different classes of allowed transformations. We begin with rigid two dimensional problems, using linear edge fragments. Details of the development are deferred to an appendix. Note that this is a specific case of using the generalized Hough transform. We will extend the arguments in later sections to deal with three-dimensional problems and to deal with problems involving change of scale.

3.1 Rigid transformations

Suppose we are considering the recognition of a two-dimensional polygonal model from noisy, occluded data. If $\mathcal{M}$ is the model coordinate system, we let

$\mathbf{M}_J$ be the vector to the midpoint of a model edge, measured in $\mathcal{M}$,

$\hat{\mathbf{T}}_J$ be the unit tangent of the edge, measured in $\mathcal{M}$,

L_J be the length of the edge.

We let $\mathbf{m}_j, \hat{\mathbf{t}}_j, \ell_j$ denote similar parameters for a data edge, measured in the sensor based coordinate system, $\mathcal{I}$. (Note that we use upper case characters to distinguish model parameters and lower case characters to distinguish sensory data parameters.)

The transformation from model coordinates to sensor coordinates may be represented by

$$\mathbf{v}_s = R_\theta \mathbf{V}_M + \mathbf{V}_0$$

where $\mathbf{V}_M$ is a vector in model coordinates, R_θ is a rotation matrix corresponding to an angle of θ, $\mathbf{V}_0$ is a translation offset, and $\mathbf{v}_s$ is the corresponding vector in sensor coordinates.

We need to know what transformations will map a model edge to a data edge. First, if $\ell_j > L_J$, we assume that the two edges cannot match (we consider the case of variable scale in the next section). Thus, suppose that $\ell_j \leq L_J$. Then the rotation needed to align the two tangents is given by the angle θ_m between $\hat{\mathbf{T}}_J$ and

$\hat{\mathbf{t}}_j$, and this defines a rotation matrix R_{θ_m}. If we apply this rotation to the set of edge points

$$\left\{ \mathbf{M}_J + \alpha \hat{\mathbf{T}}_J \quad \middle| \quad \alpha \in \left[-\frac{L_J}{2}, \frac{L_J}{2} \right] \right\}$$

we get a set of transformed points

$$\left\{ R_{\theta_m} \left[\mathbf{M}_J + \alpha \hat{\mathbf{T}}_J \right] \quad \middle| \quad \alpha \in \left[-\frac{L_J}{2}, \frac{L_J}{2} \right] \right\}.$$

To align the edges, we need to translate these rotated points. Now, because $\ell_j \leq L_J$, there are many transformations that will cause the edges to overlap. Consider one endpoint of the data edge

$$\mathbf{p}_1 = \mathbf{m}_j - \frac{\ell_j}{2} \hat{\mathbf{t}}_j.$$

If this happens to coincide with a model edge endpoint,

$$\mathbf{P}_1 = \mathbf{M}_J - \frac{L_J}{2} \hat{\mathbf{T}}_J$$

then

$$\mathbf{m}_j - \frac{\ell_j}{2} \hat{\mathbf{t}}_j = R_{\theta_m} \left[\mathbf{M}_J - \frac{L_J}{2} \hat{\mathbf{T}}_J \right] + \mathbf{V}_0$$

so that the translation is

$$\mathbf{V}_0 = \mathbf{m}_j - R_{\theta_m} \mathbf{M}_J + \frac{L_J - \ell_j}{2} R_{\theta_m} \hat{\mathbf{T}}_J$$

because $R_{\theta_m} \hat{\mathbf{T}}_J = \hat{\mathbf{t}}_j$. Similarly, if the other endpoints align, we get

$$\mathbf{V}_0 = \mathbf{m}_j - R_{\theta_m} \mathbf{M}_J - \frac{L_J - \ell_j}{2} R_{\theta_m} \hat{\mathbf{T}}_J.$$

Because any intermediate position is also acceptable, the set of translations consistent with matching model edge J to data edge j is given by

$$\left\{ \mathbf{m}_j - R_{\theta_m} \mathbf{M}_J + \alpha R_{\theta_m} \hat{\mathbf{T}}_J \quad \middle| \quad \alpha \in \left[-\frac{L_J - \ell_j}{2}, \frac{L_J - \ell_j}{2} \right] \right\}. \tag{1}$$

Hence, matching model edge J to data edge j yields a set of points in transform space $\mathcal{P}$, with a single value for the rotation parameter and a set of values for the translation, that correspond to a line of length $L_J - \ell_j$, with orientation $R_{\theta_m} \hat{\mathbf{T}}_J$ in the x–y plane.

This, however, ignores the issue of noise in the measurements. In practice, we may only know the position of the endpoints of the data edge to within some ball (which in two dimensions is just a circle) of radius ϵ_p, and the orientation to within an angular error of ϵ_a. For the case of two dimensional lines, these error ranges are related. Given endpoint variations of ϵ_p, it is straightforward to show that the maximum angular variation is when the correct line is tangent to both circles of radius ϵ_p about the two endpoints, and is given by

$$\epsilon_a = \tan^{-1} \left(\frac{2\epsilon_p}{\sqrt{\ell^2 - 4\epsilon_p^2}} \right)$$

provided $\ell > 2\epsilon_p$.

Inclusion of error effects on position measurements imply that the line of feasible translations, for a given rotation, (as given by equation (1)), must be expanded to include any points in the parameter space within ϵ_p of that line. Further, this expansion into a region must be repeated for each value of θ in $[\theta_m - \epsilon_a, \theta_m + \epsilon_a]$. Note that this carves out a skewed volume in Hough or transform space, because the region's center and orientation are functions of θ (see equation (1)). This observation has been carefully analyzed in [Clemens 86].

Thus, given $\mathbf{M}_J, \hat{\mathbf{T}}_J, L_J, \mathbf{m}_j, \hat{\mathbf{t}}_j, \ell_j$, we will use the following conditions:

- If $\ell_j - 2\epsilon_p > L_J$, then there are no consistent transformations,
- Otherwise, the set of feasible transformations is denote by the volume

$$\mathcal{V}(j,J) = \bigcup_{\theta \in [\theta_m - \epsilon_a, \theta_m + \epsilon_a]} \mathcal{S}(\theta, j, J)$$

where an individual set of translations is denoted by:

$$\mathcal{S}(\theta,j,J) = \left\{ (\theta, \mathbf{V}_0) \middle| \exists \alpha, |\alpha| \leq \frac{L_J - \ell_j}{2}, \|\mathbf{m}_j - R_\theta \mathbf{M}_J + \alpha \hat{\mathbf{T}}_J - \mathbf{V}_0\| \leq \epsilon_p \right\}.$$

These conditions imply that if a model-data pair of edges satisfy the unary constraint of length agreement, then there is a set of transforms that must all be considered as consistent.

We can already use these results to estimate the size of the set of feasible transformations. Some simple manipulations indicate that the volume of the region defined above is given by

$$2\epsilon_a \left[2\epsilon_p (L_J - \ell_j) + \pi \epsilon_p^2 \right].$$

Of more interest is the number of buckets in the Hough space that are consistent with such volumes.

If the Hough space $\mathcal{H}$ were continuous, and hence identical to the transform space $\mathcal{P}$, then we would simply need to compute all such volumes, over all data-model pairings, and let $f(\theta, \mathbf{t})$ denote the number of volumes that contain the point $(\theta, \mathbf{t})$. Then the correct interpretation would be the point at which f attains a maximum. However, in real systems, one usually tessellates the transform space $\mathcal{P}$ into non-infinitesimal buckets to obtain the Hough space $\mathcal{H}$. We let the dimensions of the Hough buckets be h_θ along the rotation axis, and h_t along each of the translation axes. Thus, we really want to determine the number of buckets that intersect one of these volumes, as that will determine the redundancy of the hashing scheme.

We begin by considering the plane of buckets consistent with a rotation value of θ_c. Suppose we let $B(\theta, j, J)$ denote the set of buckets in this plane that intersect the slice $\mathcal{S}(\theta, j, J)$. As θ varies from θ_c to $\theta_c + h_\theta$, the slice $\mathcal{S}(\theta, j, J)$ changes, and hence the set $B(\theta, j, J)$ may also change. To determine the entire set of buckets

$$\bigcup_{\theta \in [\theta_c, \theta_c + h_\theta]} B(\theta, j, J)$$

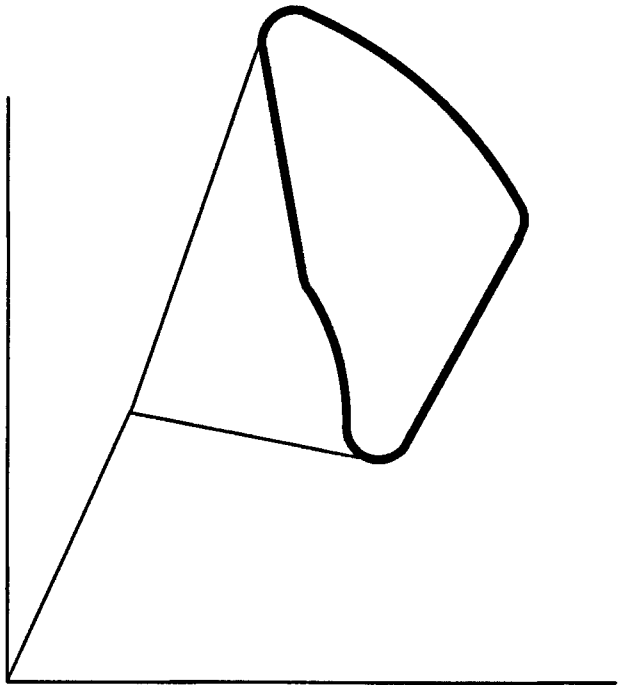

Figure 1. Region of feasible translations. The outlined area denotes the set of translations that are consistent with a data-model pairing, as the orientation ranges over the size of a Hough bucket. Details of the development are given in the appendix.

we can first project each slice in x–y, $\mathcal{S}(\theta, j, J)$ onto the x–y plane, and then find the number of buckets that intersect the union of these projections.

The set of feasible translations under this projection is shown in Figure 1. In the appendix, we show that a lower bound on the expected redundancy factor for pose clustering, b, i.e. the number of buckets into which a single data-model pairing casts a vote is given by

$$b \geq \left\lceil \frac{2\epsilon_a}{h_\theta} \right\rceil \left\lceil A_+^*(h_\theta, \epsilon_p^*, M^*, L^*, \beta) \right\rceil \tag{2a}$$

where the bound on angular error is given by

$$\epsilon_a = \tan^{-1}\left(\frac{2\epsilon_p^*}{\sqrt{(\ell^*)^2 - 4(\epsilon_p^*)^2}} \right) \tag{2b}$$

and where the modified area is given by

$$A_+^*(h_\theta, \epsilon_p^*, M^*, L^*, \beta) \geq 2M^*(1-\beta)L^*\frac{h_\theta}{\pi} + \pi\left(\epsilon_p^*\right)^2 + 2\epsilon_p^* L^*(1-\beta) + 2\epsilon_p^* h_\theta M^* + \frac{1}{2\sqrt{2}}\left(2M^* h_\theta + 2(1-\beta)L^* + 2\pi\epsilon_p^*\right). \tag{2c}$$

Note that this expression depends on the distance of the midpoint of the model edge from the center of the coordinate system, M, on the model edge length L,

on the length of the data edge, which we have assumed to be a fraction βL of the model edge length, on the bound in position uncertainty, ϵ_p, and on the size of the rotational dimension of the Hough bucket. The expressions also depend on the size of the translation dimension of the Hough bucket, which we have normalized for, using

$$\epsilon_p^* = \frac{\epsilon_p}{h_t} \qquad L^* = \frac{L}{h_t} \qquad M^* = \frac{M}{h_t}.$$

We have omitted the subscripts in the above expression, in an attempt to maintain readability of the expression.

3.1.1 Examples

To demonstrate the effect of this redundancy, we consider some representative examples. For simplicity, we will consider an object with equal length sides ($L_J = L = 50$ pixels $\forall J$) and with constant offset of the midpoint of each edge from the centroid of the object, ($M_J = M = 100$ pixels $\forall J$). We will assume that the size of the image is 500 pixels on a side. We consider two different tessellations of the Hough space, $h_t = 5, h_\theta = \pi/36$ and $h_t = 25, h_\theta = 5\pi/36$. For each of these, we consider three different error bounds on the sensory data, $\epsilon_p = 2.5, 5$ and 10 pixels. We also consider three different levels of fragmentation of the data edges, that is, the fraction of the model edge actual obtained in the image as a data edge. This is given by setting $\beta = 2\epsilon_p/L, .5, 1.0$, corresponding to the smallest allowed size, to half the size of the model edge, and to the case of no occlusion of the edges. Recall that β refers to the ratio of the length of the deta edge to the length of the model, and reflects the amount of occlusion present in an individual edge. Tables 1 and 2 summarize the redundancy b for each of these case, shown both in terms of the actual number of buckets, and as a fraction of the total buckets in the tessellated space, using the bounds of equation (2).

	$\beta = \frac{2\epsilon_p}{L}$		$\beta = .5$		$\beta = 1$	
$\epsilon_p = 2.5$	1116	.00155	95	.00013	15	.00002
5	1476	.00205	300	.00042	55	.00008
10	2196	.00305	1210	.00168	260	.00036

Table 1. Redundancy of Hough hashing, for tessellations of $h_t = 5$ and $h_\theta = \pi/36$. The lower bound on actual number of buckets hashed, and the fraction of the total number of buckets is given, for a single data-model pairing. The total number of buckets in this case is 720,000.

The redundancies reported above apply to a single data-model pairing, and the examples reported in Tables 1 and 2 use particular values of the length of the model edge, and its offset from the origin of the model coordinate system. Very similar redundancies hold for other values of these parameters, however. In Table 1a, we

	$\beta = \frac{2\epsilon_p}{L}$		$\beta = .5$		$\beta = 1$	
$\epsilon_p = 2.5$	48	.00857	4	.00071	2	.00036
5	46	.01000	10	.00179	3	.00054
10	64	.01143	35	.00625	10	.00179

Table 2. Redundancy of Hough hashing, for tessellations of $h_t = 25$ and $h_\theta = 5\pi/36$. The lower bound on actual number of buckets hashed, and the fraction of the total number of buckets is given, for a single data-model pairing. The total number of buckets in this case is 5,600.

show the redundancies obtained for fixed values of error, $\epsilon_p = 5$, and a fixed bucket size, $h_t = 5, h_\theta = \pi/36$, but with varying edge length L and varying model offset M. One can see that considerable variation in these values yields similar redundancies.

	M= 50	100	200
L=25	352	440	616
=50	250	300	400
=100	205	245	325

Table 1a. Redundancy of Hough hashing, for tessellations of $h_t = 5$ and $h_\theta = \pi/36$. The lower bound on actual number of buckets hashed is given for a single data-model pairing. The error is fixed at $\epsilon_p = 5$, occlusion is fixed at $\beta = .5$ and the length and offset of the model edges are varied. The total number of buckets in this case is 720,000.

The data in Tables 1 and 2 deal with extended edge fragments. If the data is point data, for example, vertices, then $\beta = 1$ and $L = 1$. In this case, we need some other means of estimating the orientation, and for illustrative purposes we use $\epsilon_a = \pi/36$. This is a tighter bound than that used in the previous examples. The redundancy for the two different tessellations of the Hough space, and for the different positional error bounds are shown in Table 3.

	$h_t = 5, h_\theta = \frac{\pi}{36}$		$h_t = 25, h_\theta = \frac{5\pi}{36}$	
$\epsilon_p = 2.5$	10	.00001	2	.00036
5	22	.00003	3	.00054
10	52	.00007	5	.00089

Table 3. Redundancy of Hough hashing, for point data. The error in measuring the normal is assumed to be $\epsilon_a = \pi/36$. The lower bound on actual number of buckets hashed, and the fraction of the total number of buckets is given, for a single data-model pairing. The number of buckets is 720,000 for the left part of the table, and 5,600 for the right.

All of the above examples involve the use of a full three-parameter Hough space. In many cases, it is common to use the projection of that space onto a smaller subspace,

	$h_t = 5, h_\theta = \frac{\pi}{36}$		$h_t = 25, h_\theta = \frac{5\pi}{36}$	
$\epsilon_p = 2.5$	8	.0008	1	.0025
5	15	.0015	2	.0050
10	34	.0034	3	.0075

Table 6. Redundancy of Hough hashing, for point data, using projection of the full space onto the two dimensional translation subspace. The error in measuring the normal is assumed to be $\epsilon_a = \pi/36$. The lower bound on actual number of buckets hashed, and the fraction of the total number of buckets is given, for a single data-model pairing. The number of buckets is 10,000 for the left part of the table, and 400 for the right.

typically using the projection onto the translational subspace. We can also derive estimates of the redundancy of this method. We can use the same equations as before, with some minor changes. First, the swept area of the translational subspace is given by considering the full range of rotational values, $2\epsilon_a$ in place of h_θ. Second, the redundancy factor is obtained by considering only the translational subspace, and is given by

$$b \geq \left\lceil A_+^*(2\epsilon_a, \epsilon_p^*, M^*, L^*, \beta) \right\rceil . \tag{3}$$

Examples of the redundancy, using equation (3) are shown in Tables 4, 5, and 6.

	$\beta = \frac{2\epsilon_p}{L}$		$\beta = .5$		$\beta = 1$	
$\epsilon_p = 2.5$	485	.0485	50	.0050	9	.0009
5	518	.0518	116	.0116	28	.0028
10	582	.0582	334	.0334	95	.0095

Table 4. Redundancy of Hough hashing, for tessellations of $h_t = 5$ and $h_\theta = \pi/36$, using projection of the full space onto the two dimensional translation subspace. The lower bound on actual number of buckets hashed, and the fraction of the total number of buckets is given, for a single data-model pairing. The total number of buckets in this case is 10,000.

	$\beta = \frac{2\epsilon_p}{L}$		$\beta = .5$		$\beta = 1$	
$\epsilon_p = 2.5$	28	.07	4	.01	1	.0025
5	30	.075	8	.02	3	.0075
10	32	.08	19	.0475	7	.0175

Table 5. Redundancy of Hough hashing, for tessellations of $h_t = 25$ and $h_\theta = 5\pi/36$, using projection of the full space onto the two dimensional translation subspace. The lower bound on actual number of buckets hashed, and the fraction of the total number of buckets is given, for a single data-model pairing. The total number of buckets in this case is 400.

Several observations are in order. First, the tables show that in general, the redundancy of Hough hashing can be quite large, both in terms of the number of buckets consistent with a single data-model pairing, and in terms of the fraction of the Hough space deemed consistent with such a pairing. As expected, when one considers the matching of vertices to vertices in place of matching edges to edges, the redundancy improves. This is to be expected, since in the edge case, a partial edge can slide along its corresponding model edge, leading to more feasible transformations.

As well, when the sensor error is reduced, the redundancy improves. Increasing the coarseness of the Hough tessellation can reduce the total number of buckets into which a data-model pair votes, but in general this increases the fraction of the total number of buckets selected. In general, the analysis and examples argue that the redundancy of Houghing can be quite high.

While we have provided examples of several levels of sensor error, we note that the higher levels of error are probably more indicative of the situation encountered with real images. Several factors will contribute to the bound for ϵ_p. First, aberrations in the optics will cause the recorded edges to deviate from the actual physical edge. Second, smoothing effects in the edge detector will add to the displacement of recorded edges. The amount of deviation will depend on the specifics of the operator, but 1 or 2 pixel errors are likely to be common. Third, using a split-and-merge operation to extract linear segments from grey level edges will further add to the error, typically by several pixels, so that overall error bounds of at least 5 pixels are to be expected.

3.2 Scaled transformations

Suppose we now allow the objects to scale, as well as rotate and translate. In this case the transformation from model to sensor coordinates is given by

$$\mathbf{v}_s = kR_\theta \mathbf{V}_M + \mathbf{V}_0$$

where $\mathbf{V}_M$ is a vector in model coordinates, R_θ is a rotation matrix corresponding to an angle of θ, $\mathbf{V}_0$ is a translation offset, k is a scale factor and $\mathbf{v}_s$ is the corresponding vector in sensor coordinates.

In this case, the set of feasible translations corresponding to a data-model pairing is a function of both the scale and the rotation:

$$\left\{ \mathbf{m}_j - kR_{\theta_m}\mathbf{M}_J + \alpha R_{\theta_m}\hat{\mathbf{T}}_J \;\middle|\; \alpha \in \left[-\frac{kL_J - \ell_j}{2}, \frac{kL_J - \ell_j}{2} \right] \right\}.$$

In this case, the scale has a minimum bound of

$$k \geq \frac{\ell_j}{L_J}.$$

To determine the redundancy factor for parameter hashing in the case of scale, we again want to determine the number of buckets consistent with a data-model pairing for a single slice of the x–y components of the transform space. Note that in this case, the transform space $\mathcal{T}$ is four dimensional, with an extra axis for the

scale factor. Projecting the volume obtained as θ varies over the bounds of a single bucket gives us the volume shown in Figure 2, where now the borders are functions of the scale factor. If we now look at the projection of the volume as k is varied, we will get the region obtained by varying the region in Figure 5 over the range of values of k. This new region is shown in Figure 3.

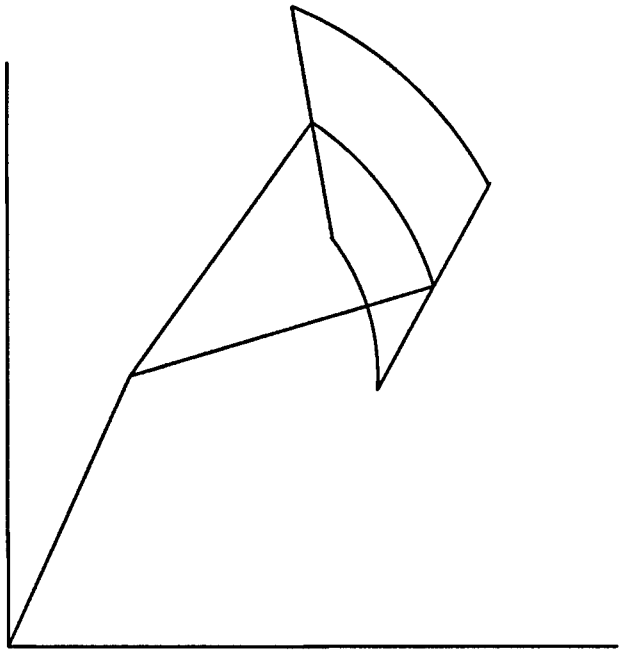

Figure 2. Rotation of the line of feasible translations through h_θ radians.

Using an analysis similar to the previous case (details are given in the appendix), we can derive bounds on the redundancy in the case of objects that can scale. Suppose we define the full range of possible scale factors to be $[1, k_{max}]$, so that the model is defined as the smallest possible instance of an object. Then to count the redundancy factor in this case, we must sum the number of buckets obtained over all possible scale factors. If the spacing of the Hough buckets in the scale dimension is h_k, then this sum is given by:

$$b_s \geq \left\lceil \frac{2\epsilon_a}{h_\theta} \right\rceil \left\lceil A^*_{s,+}(i_s - \max\{1, \frac{\ell^*}{L^* h_k}\}, i_s h_k) \right\rceil + \sum_{i=i_s}^{\frac{k_{max}}{h_k}} \left\lceil \frac{2\epsilon_a}{h_\theta} \right\rceil \left\lceil A^*_{s,+}(h_k, i h_k) \right\rceil \quad (5)$$

where

$$i_s = \left\lceil \frac{\max(1, \frac{\ell^*}{L})}{h_k} \right\rceil$$

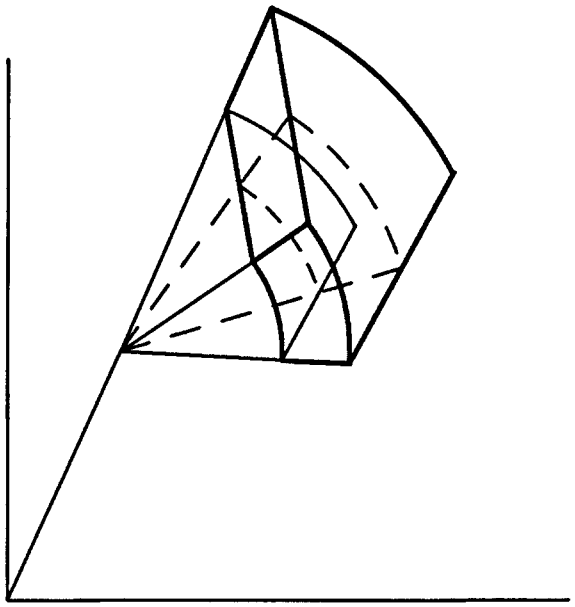

Figure 3. Region of translation space consistent with scale variation and angle variation.

and where

$$A^*_{s,+} = A^*_s + \frac{P^*_s}{2\sqrt{2}}$$

$$P^*_s \geq h_\theta \left((k_h + k_\ell) M^*_J - \frac{k_h - k_\ell}{2} L^*_J \right) + (k_h + k_\ell) L^*_J - 2\ell^*_j + 2(k_h - k_\ell) M^*_J + 2\pi\epsilon^*_p$$

$$\begin{aligned} A^*_s(\Delta k, k_h) =& h_\theta k_h M^*_J \left(k_h L^*_J - \ell^*_j \right) \frac{2}{\pi} \\ &+ h_\theta \Delta k \left[\left(k_h - \frac{\Delta k}{2} \right) \left((M^*_J)^2 + \frac{(L^*_J)^2}{4} \right) - \frac{L^*_J \ell^*_j}{2} \right] \\ &+ \frac{h_\theta}{\pi} M^*_J \left[(2k_h^2 + 2k_h \Delta k + \Delta k^2) L^*_j - (2k_h + \Delta k) \ell^*_j \right] \\ &+ M^*_J \Delta k \left[\left(k_h - \frac{\Delta k}{2} \right) L^*_J - \frac{\ell^*_j}{2} \right] \frac{2}{\pi} \\ &+ h_\theta \epsilon^*_p \left[(2k_h - \Delta k) M^*_J - \frac{\Delta k}{2} L^*_J \right] + 2\epsilon^*_p \left(2 \left(k_h - \frac{\Delta k}{2} \right) L^*_J - 2\ell^*_j \right) \\ &+ 2\epsilon^*_p \Delta k M^*_J + \pi (\epsilon^*_p)^2. \end{aligned}$$

The final, rather messy, expression is a function of the range of variation in scale Δk as well as the maximum value of the scale parameter k_h.

We can use this to generate example redundancies. Tables 7 and 8 show the redundancy, for the case of $M = 100, L = 50$ using a fine tessellation of $h_t = 5, h_\theta = \frac{\pi}{36}$ and using 100 buckets in the scale dimension. We consider both the case of $k_{max} = 2$ and $k_{max} = 10$ (Tables 7 and 8 respectively).

	$\beta = .5$		$\beta = 1$		$\beta = 1.5$	
$\epsilon_p = 2.5$	36320	.00050	12444	.00017	2742	.00004
$\epsilon_p = 5$	99190	.00138	28925	.00040	7872	.00011
$\epsilon_p = 10$	345752	.00480	95500	.00133	23863	.00033

Table 7. Redundancy of Hough hashing, including a scale dimension with range from 1 to 2 in increments of .01. Tessellations are $h_t = 5$ and $h_\theta = \pi/36$. The lower bound on expected number of buckets hashed, and the fraction of the total number of buckets is given, for a single data-model pairing. The total number of buckets in this case is $72,000,000$.

	$\beta = .5$		$\beta = 1$		$\beta = 5$	
$\epsilon_p = 2.5$	518455	.00720	286998	.00399	37887	.00053
$\epsilon_p = 5$	1153870	.01603	531745	.00739	41820	.00058
$\epsilon_p = 10$	3063967	.04256	1281950	.01780	99908	.00139

Table 8. Redundancy of Hough hashing, including a scale dimension with range from 1 to 10 in increments of .09. Tessellations are $h_t = 5$ and $h_\theta = \pi/36$. The lower bound on the expected number of buckets hashed, and the fraction of the total number of buckets is given, for a single data-model pairing. The total number of buckets in this case is $72,000,000$.

We can also do the parameter hashing by projecting onto a subspace of the full space. In the case of allowing scale to vary, for instance, we can consider the projection of the 4D volume into the normal 3D space spanned by the two translational and one rotational dimensions. The data for the cases of Tables 7 and 8 under this projection are given in Tables 9 and 10.

	$\beta = .5$		$\beta = 1$		$\beta = 1.5$	
$\epsilon_p = 2.5$	1860	.00258	897	.00125	310	.00043
$\epsilon_p = 5$	4180	.00581	1675	.00233	704	.00098
$\epsilon_p = 10$	11308	.01571	4120	.00572	1568	.00218

Table 9. Redundancy of Hough hashing, including a scale dimension with range from 1 to 2, projected onto the normal 3D space. Tessellations are $h_t = 5$ and $h_\theta = \pi/36$. The lower bound on the expected number of buckets hashed, and the fraction of the total number of buckets is given, for a single data-model pairing. The total number of buckets in this case is $720,000$.

All of the examples of this section argue strongly that the redundancy of Hough transforms, in the presence of sensor error and partial occlusion of data elements, is quite high. In particular, the number of buckets in the Hough space that are consistent with a data-model pairing can be a significant portion of the total Hough

	$\beta = .5$		$\beta = 1$		$\beta = 5$	
$\epsilon_p = 2.5$	59130	.08213	34113	.04738	6393	.00888
$\epsilon_p = 5$	121180	.16831	58260	.08092	6558	.00911
$\epsilon_p = 10$	279488	.38818	122190	.16971	13788	.01915

Table 10. Redundancy of Hough hashing, including a scale dimension with range from 1 to 10, projected onto the normal 3D space. Tessellations are $h_t = 5$ and $h_\theta = \pi/36$. The lower bound on the expected number of buckets hashed, and the fraction of the total number of buckets is given, for a single data-model pairing. The total number of buckets in this case is 720,000.

space. This relative redundancy increases with increasing error, with occlusion of data edges, when scaling is included as a free parameter, and when projections of the full parameter space onto subspaces is used. When point data are used with minimal error, the redundancy of the Hough technique is more reasonable, but in general cases, the method has severe redundancy problems.

3.2 Three dimensional problems

We can also extend our method of analysis to three dimensional problems. In this case, we assume that we are matching planar patches of 3D data, together with an estimate of the surface normal of the patch, against comparable planar model faces. For ease of analysis, we will assume circular patches. As in the 2D case, we need to determine the volume in transform space consistent with a pairing of a data patch and a model face, and then determine the number of Hough buckets intersected by the volume.

To represent the transform space, we use:

- a cubic cell tesselation of the subset of $\mathcal{R}^3$ defining legitimate translations of the mode. Each bucket has sides of size h_t.
- a partition of the surface of the Gaussian sphere, used to denote the axis of rotation of the model. Each section has an area of h_r.
- a partition of the range $[0, 2\pi)$ for the angle of rotation about the axis given above. Each section has a size of h_θ.

Now, we first consider the rotation part of the transform. Given a model normal $\hat{\mathbf{N}}$ and a measured data normal $\hat{\mathbf{n}}$, there is a set of rotation vectors, and associated angles, that will cause $\hat{\mathbf{N}}$ to rotate into $\hat{\mathbf{n}}$. This set of rotation vectors $\{\hat{\mathbf{r}}\}$ consists of those unit vectors lying on the great circle of points on the Gaussian sphere, equidistant from $\hat{\mathbf{n}}$ and $\hat{\mathbf{N}}$. Equivalently, they are the set of unit vectors $\hat{\mathbf{r}}$ such that

$$< \hat{\mathbf{r}}, \hat{\mathbf{N}} - \hat{\mathbf{n}} >= 0$$

where the special case of $\hat{\mathbf{N}} = \hat{\mathbf{n}}$ is treated separately.

Now, the data normal $\hat{\mathbf{n}}$ is not exact, but deviates from the correct normal by some error. We assume that $\hat{\mathbf{n}}$ lies within a bounded range of the actual normal $\hat{\mathbf{n}}_0$,

given by

$$< \hat{\mathbf{n}}, \hat{\mathbf{n}}_0 > \geq \cos \epsilon_a.$$

We need to estimate the set of feasible rotation vectors $\hat{\mathbf{r}}$ as $\hat{\mathbf{n}}$ varies over the ϵ_a-cone about $\hat{\mathbf{n}}_0$. That set is given by the region swept out on the Gaussian sphere by the great circle perpendicular to the unit vector $\hat{\mathbf{m}}$ in the direction of $\hat{\mathbf{N}} - \hat{\mathbf{n}}$ as $\hat{\mathbf{n}}$ varies over the range defined by

$$< \hat{\mathbf{n}}, \hat{\mathbf{n}}_0 > \geq \cos \epsilon_a.$$

To see what this looks like, consider the case in which $\hat{\mathbf{n}}_0 = -\hat{\mathbf{N}}$. Then $\hat{\mathbf{m}}$ must lie in an $\frac{\epsilon_a}{2}$-cone about $\hat{\mathbf{N}}$ (since $\hat{\mathbf{N}} - \hat{\mathbf{n}}_0 = 2\hat{\mathbf{N}}$ and $\hat{\mathbf{n}}$ varies within a cone spanned by ϵ_a, the intersection of this cone with the Gaussian sphere gives an $\frac{\epsilon_a}{2}$ cone). This means that the perpendicular great circle sweeps out a band about the great circle perpendicular to $\hat{\mathbf{N}}$, with a maximum deviation of $\frac{\epsilon_a}{2}$ on either side. We can straightforwardly evaluate the area swept out, and it is given by

$$4\pi \sin \frac{\epsilon_a}{2}.$$

Now consider what happens as $\hat{\mathbf{n}}_0$ varies from the special case of $\hat{\mathbf{n}}_0 = -\hat{\mathbf{N}}$. We let α denote the angle between $\hat{\mathbf{N}}$ and $\hat{\mathbf{n}}_0$. First, the length of the vector $\hat{\mathbf{N}} - \hat{\mathbf{n}}_0$ decreases to

$$2 \sin \frac{\alpha}{2}.$$

Second, the ϵ_a-cone about $\hat{\mathbf{n}}_0$ now becomes a skewed cone about $\hat{\mathbf{N}} - \hat{\mathbf{n}}_0$. We can get a lower bound on the size of the largest regular cone contained within this skewed cone. The geometry is shown in Figure4.

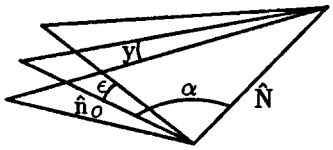

Figure 4. Geometry for determining the cone of possible vectors $\hat{\mathbf{N}} - \hat{\mathbf{n}}$ for $< \hat{\mathbf{n}}, \hat{\mathbf{n}}_0 > \geq \cos \epsilon$.

To determine the scope of this new cone, we need to solve for y, as shown in the figure. Appropriate trigonometry yields

$$\tan y = \frac{\sin \epsilon \tan \frac{\alpha}{2}}{2 \tan \frac{\alpha}{2} + \sin \epsilon}.$$

As $\hat{\mathrm{N}} - \hat{\mathrm{n}}$ varies over this cone, the great circle perpendicular to it will sweep out an area on the surface of the Gaussian sphere, and simple integration shows that this area is given by

$$\frac{4\pi}{\sqrt{1+\left(\frac{2}{\sin\epsilon}+\frac{1}{\tan\frac{\alpha}{2}}\right)}}.$$

We need to obtain a bound for this area. This expression is minimized for $\alpha = 0$, but this corresponds to the special case in which $\hat{\mathrm{N}}$ is a fixed point of the rotation. In this case, while there is no uncertainty in the axis of rotation, there is complete uncertainty in the angle of rotation, and hence in this case, such a pairing would intersect

$$\frac{2\pi}{h_\theta}$$

buckets in the rotation part of the transform space. In general, however, the surface normal will not be a fixed point of the rotation. In this case, (which we treat as $\alpha \geq \epsilon$ to handle the noise in the system), the rotation angle is uniquely determined, but the axis of rotation is uncertain. The minimum uncertainty is given by $\alpha = \epsilon_a$, and the minimum area swept out on the Gaussian sphere is bounded below by

$$\frac{4\pi}{\sqrt{1+\left(\frac{2}{\sin\epsilon}+\frac{1}{\tan\frac{\epsilon}{2}}\right)^2}}.$$

Hence, given that each bucket in the Hough space has an area on the Gaussian sphere of h_r, a pairing of a model and data patch intersects at least

$$\frac{4\pi}{\sqrt{1+\left(\frac{2}{\sin\epsilon}+\frac{1}{\tan\frac{\epsilon}{2}}\right)^2}}\frac{1}{h_r}.$$

Next, we consider the translation component of the transform. Suppose we have a model patch of radius R and a data patch of radius r. Once we have rotated the model, we can slide the transformed model patch so that it contains the data patch. There are a set of possible translations consistent with this, and they are delimited by a circle of radius $R - r$ in some slice of the translation components of the transform space. When we include the effects of positional error (ϵ_p), we get a disk of radius $R - r + \epsilon_p$ and height ϵ_p, so that the volume of consistent translations is

$$\pi(R - r + \epsilon_p)^2 2\epsilon_p$$

and hence such a volume intersects at least

$$\frac{2\pi\epsilon_p(R - r + \epsilon_p)^2}{h_t^3}$$

buckets. Thus, by putting all of this together, we see that the redundancy factor in the 3D case is bounded by:

$$b \geq \left\lceil \frac{4\pi}{\sqrt{1+\left(\frac{2}{\sin\epsilon}+\frac{1}{\tan\frac{\epsilon}{2}}\right)^2}}\frac{1}{h_r} \right\rceil \left\lceil \frac{2\pi\epsilon_p(R - r + \epsilon_p)^2}{h_t^3} \right\rceil.$$

As an example, we consider the case in which $\epsilon_a = \pi/10, \epsilon_p = 5, h_\theta = 4\pi/100, h_t = 5$, the results of which are shown in Table 11.

	$\beta = .2$	.5	1.0
R=20	888	456	56
=50	4072	1816	56
=100	14528	6088	56

Table 11. Redundancy of Hough hashing, for three dimensional problems. Tesselations are $h_t = 5$ and $h_r = 4\pi/200$. The lower bound on actual number of buckets is given, for a single data-model pairing. The parameters varied are the amount of occlusion β and the size of the model face R.

Note that the bounds derived here are quite weak. We could obtain much tighter bounds, but feel that these suffice to demonstrate that the same problems observed in two dimensions also hold in three.

4. An Occupancy Model of the Hough Transform

The previous section has addressed the issue of the number of Hough buckets that are consistent with a pairing of a sensory feature and a model feature. The second question to be addressed in considering the efficacy of the Hough transform for finding solutions to the recognition problem, is the likelihood of large random clusters occurring at random in Hough space.

Recall that the recognition problem, when using Hough transforms, is to use all pairings of model and image features to compute transformations from the model to the image. Each parameter of a given transformation is quantized, and the transformation is entered into the appropriate buckets of an n-dimensional table. Buckets containing a large number of transformations (a peak) are taken to correspond to an instance of the object in the image. Significantly large clusters are either identified by a threshold on the number of transformations in a bucket, or by using the largest few buckets. In either case, the size of peak, l, that corresponds to a correct match of the model to the image should be large enough that it is not likely to occur at random. Note that l will be at most some fraction of m, corresponding to the fraction of the model features that are matched to image features.

In this section, we consider the robustness of this approach, given the bounds derived in the previous section on the number of buckets for which a single data-model pairing may vote. We model the generalized Hough transform as an occupancy problem, in order to obtain an estimate of the probability that a Hough bucket will have peaks of size l or more at random. This probability should be very small in order for

the technique to identify primarily true instances of an object in an image, rather than random groupings of features.

If the transformations from a model to an image were uniformly randomly distributed over the parameter space, then the probability that a given transformation would fall into a particular bucket would be $\frac{1}{n}$, where n is the number of buckets. If each instance was independent of the other instances, the probability that r transformations would fall into a given bucket is n^{-r}. To the extent that transformations are not uniformly and independently distributed, they will tend to clump together more than indicated by this model. Thus modeling the transformations as uniformly randomly distributed yields a conservative model of the actual distribution. The true distribution will yield random peaks that are at least as large as the uniform case.

Given a distribution of r events into n cells, one can speak of the occupancy numbers, or the number of events in each cell, denoted by $r_1, \ldots, r_n$, where each $r_i \geq 0$ and $\sum r_i = r$. If the events are randomly distributed such that each of the n^r placements have the equal probability, n^{-r}, then the probability of a given arrangement with occupancy numbers $r_1, \ldots, r_n$ is

$$p_{r_1,\ldots,r_n} = \frac{r!}{r_1! r_2! \cdots r_n!} n^{-r}.$$

This distribution of events is often termed the classical occupancy problem, or Maxwell-Boltzmann statistics (for a standard text see [Feller 68]).

For the classical occupancy problem, the probability, p_k, that a given cell contains exactly k events is given by the binomial distribution,

$$p_k = \binom{r}{k} \frac{1}{n^k} \left(1 - \frac{1}{n}\right)^{r-k}.$$

We are interested in the probability that a given cell will contain l or more events at random, which is

$$p_{\geq l} = 1 - \sum_{k=0}^{l-1} p_k.$$

The expected number of cells in a Hough table that will contain peaks of size at least l is then given by

$$E_{\geq l} = n p_{\geq l}$$

where n is the number of cells in the table. Ideally, the peaks corresponding to correct matches should be of a sufficient size, l, that $E_{\geq l} < 1$. In other words, ideally the expectation should be that there will be less than one false peak in the table.

For even moderate values of n and r, the computation of p_k becomes unwieldy. For sufficiently large values of n, however, the Poisson approximation to the binomial can be used. The error of this approximation is proportional to n^{-1}, so for n's of the size discussed in the previous subsection (10^4 or larger) the error is relatively small. Using this approximation,

$$p_k \approx \frac{\lambda^k}{k!} e^{-\lambda},$$

where $\lambda = \frac{r}{n}$. Thus the parameter λ is the ratio of the number of elements entered into the table, over the number of buckets.

In addition to the Maxwell-Boltzmann distribution, another common distribution used in occupancy problems is the Bose-Einstein statistic. This distribution has an experimental basis in particle physics, and assigns an equal probability to each of the occupancy numbers, $r_1, \ldots, r_n$. Under the Bose-Einstein model, for large r and n, the limiting case is the so-called geometric distribution, where

$$p_k \approx \frac{\lambda^k}{(1+\lambda)^{k+1}}.$$

This distribution has a long tail as $k \to \infty$, and thus predicts large peaks with a higher probability than does the Maxwell-Boltzmann model. Hence we use the more conservative model given by the Maxwell-Boltzmann distribution.

4.1 Evaluating the Generalized Hough Transform

To judge the effectiveness of the generalized Hough transform as a clustering technique, the occupancy model will be used on some representative problems. First we will use the redundancy factors obtained in Section 3 to consider some two-dimensional recognition problems. Then we will examine some empirical data from a three-dimensional recognition system.

The λ parameter of the occupancy model is the ratio of the number of events entered into the table to the number of buckets. The number of events, $r = msb$, where m is the number of model features, s is the number of sensory features, and b is the redundancy factor. Thus $\lambda = msb/n$, where n is the number of buckets in the table.

We are interested in the likelihood of random peaks that are at least as large as those due to a correct match, where l is the size peak that is expected to result from a correct match. A match that correctly pairs all the model with image features will result in a peak of size $l = m$. Thus in general $l = fm$, where $0 < f \leq 1$ is the proportion of model features that are correctly matched to image features. For a given problem, the values of b and n are fixed, and we will vary m and s to determine how many peaks of size l will occur at random, for $l = .5m$, $l = .75m$, and $l = .9m$.

First we consider the case of using just the two translation parameters to enter transformations into the Hough table. With 5 pixel buckets there are a total of $n = 10,000$ buckets. If the features are edges, then each pair of model and image features defines a range of transformations that intersect $b = 116$ buckets (with an error range of $\epsilon_p = 5$ pixels and a fragmentation of $\beta = .5$, as shown in Table 4). In this case, the generalized Hough technique is very poor at finding clusters that are due to a correct match. If there are more than 47 sensory data points, then the expected number of peaks of size l occurring at random will be always be larger than 1, for any value of $l \leq m$. In other words, there will always be false matches if there

are more than 47 features in the image. Not only will there be more than one large peak at random, there will generally be many large peaks. For example, with 10 model edges and 100 image edges the expectation is that 7209 of the 10,000 buckets will contain peaks of size 10 or more. Thus a two-dimensional translation Hough table is not well suited to the problem of clustering transformations by matching edges, even for uncluttered images with moderate error and occlusion.

When the features consist of vertices rather than edges, the corresponding redundancy factor, b, is 15 (for an error range of $\epsilon_p = 5$, as shown in Table 6). The expected number of peaks that will occur at random are shown in Table 12, for peak sizes of $l = .5m$, $l = .75m$, and $l = .9m$. Cases where the expectation is less than 1 are indicated by a dash. Even though the number of redundant entries in the Hough table is much smaller for vertices than for edges, the number of false peaks is still quite high, even for moderately complex images. For example, for an image with 200 vertices and a model with 20 vertices, 90% of the model vertices must be matched in order for the expected number of false matches to be low (in this case 8). If only half of the model vertices are accounted for, then nearly every fourth bucket (2236 out of 10,000) will have a cluster as large as that resulting from a correct match.

	$f = .5$	.75	.9
$s = 200$, $m = 50$	960	1	-
$m = 20$	2236	103	8
$m = 10$	3225	863	148
$m = 5$	5591	2895	1211
$s = 100$, $m = 20$	11	-	-
$m = 10$	186	9	-
$m = 5$	1734	405	73

Table 12. Expected number of peaks occurring at random for various numbers of sensory features, s, model features, m, and visible fractions of model features, f ("-" indicates a value of < 1). For vertex features, where $b = 15$, and with a Hough table of $n = 10,000$ buckets.

The more model features that are correctly matched to image features, the larger the resulting cluster of transformations. Thus, another means of quantifying the power of the generalized Hough technique is to consider what the minimum number of model features must be in order for there to be an expectation of less than one random peak of size $l = fm$ in the Hough table. This value is shown in Table 13 for the task just considered, of a 10,000 bucket Hough table, vertex features, and $b = 15$. The entry N.P. for $s = 250$ and $f = .5$ means that there is no possible model size such that the expected number of peaks of size $.5m$ is less than 1 when there

	$f = .5$	.75	.9
$s = 250$	N.P.	52	30
100	30	14	9

Table 13. Size model required to have an expectation of less than one random cluster at least as big as the correct match, for various numbers of sensory features, s, and visible fractions of model features, f. For vertex features, where $b = 15$, and with a Hough table of $n = 10,000$ buckets.

are 250 or more image vertices. Thus again we see the limitation of this clustering method for the recognition of moderately cluttered scenes.

Next we consider the case of using all three parameters to perform the clustering. For translation buckets of 5 pixels and rotation buckets of $\pi/36$ radians, there are a total of $n = 720,000$ buckets. For edge features, the redundancy factor, b, is 300 (with an error range of $\epsilon_p = 5$ pixels and a fragmentation of $\beta = .5$, as shown in Table 1). The expected number of peaks occurring at random are shown in Table 14, for peak sizes of $l = .5m$, $l = .75m$, and $l = .9m$. For a moderately cluttered image, with $s = 500$ edges, and a model with $m = 10$ edges, the expected number of false peaks is over 40,000 if only half of the model edges are matched to image edges. If 9 of the 10 model edges are matched, then there is still an expectation of 229 false peaks.

	$f = .5$	.75	.9
$s = 1000$, $m = 100$	82,383	2	-
$m = 50$	149,009	625	2
$m = 25$	253,053	14,703	840
$m = 10$	290,655	97,702	19,260
$s = 500$, $m = 50$	63	-	-
$m = 25$	5326	7	-
$m = 10$	43,549	4048	229
$s = 250$, $m = 25$	13	-	-
$m = 10$	1008	15	-
$s = 100$, $m = 10$	53	-	-

Table 14. Expected number of peaks occurring at random for various numbers of sensory features, s, model features, m, and visible fractions of model features, f ("-" indicates a value of < 1). For edge features, where $b = 300$, and with a Hough table of $n = 720,000$ buckets.

Table 15 shows the number of model features required in order for there to be an

expectation of less than one random peak of size $l = fm$ in the Hough table. For a relatively cluttered image with 500 edges, and high occlusion of 50%, a model must have at least 80 features before the expected number of false peaks is less than 1. Even for a simple image with only 100 edges, with moderate occlusion of 25%, a model must have at least 10 edges for there to be an expectation of no false matches. Thus even using the full three parameters for clustering, there is a high likelihood that random clusters will be as large as those due to a correct match.

	$f = .5$	.75	.9
$s = 1000$	400	104	56
500	80	28	19
250	30	16	12
100	16	10	7

Table 15. Size model required to have an expectation of less than one random cluster at least as big as the correct match, for various numbers of sensory features, s, and visible fractions of model features, f. For edge features, where $b = 300$, and with a Hough table of $n = 720,000$ buckets.

Finally, we consider the case of using vertex features and the three-dimensional parameter Hough table with $n = 720,000$ buckets. The relevant redundancy factor is $b = 22$ (with an error range of $\epsilon_p = 5$ pixels, as shown in Table 3). Table 16 shows the expected number of peaks of a given size that will occur at random, and Table 17 shows the size model necessary to limit the expected number of false peaks to less than one. In Table 17 it can be seen that for all but very complex images, a match of a model with 10 or fewer features will result in an expectation of less than one false peak in the Hough table. Thus the method works relatively well for this case. The cost is quite high, however, because there are about two orders of magnitude more buckets to be searched than there are distinct transformations from the model to the image. The number of transformations is ms, which is at most a few thousand, whereas there are 720,000 buckets.

The Generalized Hough Method for 3D Recognition

In this section, we use some empirical data on the number of transformations from a model to an image to evaluate the power of the generalized Hough transform in a three-dimensional recognition task. As with the above results based on the analytic formulation of the two-dimensional problem, we find that the likelihood of large peaks occurring at random is very high for even moderately complex images and levels of uncertainty.

For 3D recognition, the size of a full Hough table becomes prohibitive, so only a subset of the transformation parameters are used to form the table. For example, in

	$f = .5$	.75	.9
$s = 1000$, $m = 10$	12	-	-
$m = 5$	7594	382	14
$s = 500$, $m = 5$	1996	51	-
$s = 250$, $m = 5$	511	6	-
$s = 100$, $m = 5$	83	-	-

Table 16. Expected number of peaks occurring at random for various numbers of sensory features, s, model features, m, and visible fractions of model features, f ("-" indicates a value of < 1). For vertex features, where $b = 22$, and with a Hough table of $n = 720,000$ buckets.

	$f = .5$	.75	.9
$s = 1000$	14	8	7
500	10	7	5
250	8	6	5
100	6	4	4

Table 17. Size model required to have an expectation of less than one random cluster at least as big as the correct match, for various numbers of sensory features, s, and visible fractions of model features, f. For vertex features, where $b = 22$, and with a Hough table of $n = 720,000$ buckets.

[Thompson and Mundy 87] the two parameters of rotation out of the viewing plane are used for an initial clustering. The Hough buckets are of size 2°, yielding a total of $n = 32,400$ buckets. An error range of 15° is allowed, so each transformation is entered into an average of $8^2 = 64$ buckets. A model has about $m = 5$ features, an image has about $i = 3000$ features, and this results in about 20,000 transformations. Thus a total of about $r = 1280000$ transformations are entered into the table, yielding a λ of about 40. In order for the expected number of false peaks in the table, $E_{\geq l}$, to be less than one, the peak size, l, must be 68. This is an order of magnitude larger than the number of model features, m. Peaks of size at least m, which is 5, will occur at random with a probability of 99%. In other words, this initial clustering eliminates virtually none of the candidates.

Following the initial clustering, a secondary clustering is performed using the third rotation parameter. This parameter is again quantized in 2° buckets, so there are a total of $n = 5,832,000$ buckets in the three-dimensional table. Each transformation is now entered in $8^3 = 512$ buckets in order to allow for 15° errors. Thus 20,000 transformations yields $r = 10,240,000$ table entries, and $\lambda = 1.8$. In order for $E_{\geq l} < 1$, the peak size, l, must be at least 11, which is a factor of two larger than the

number of model features. Peaks of size 5 occur with a probability of about 1%, so there will be nearly a hundred thousand false peaks in the three-dimensional Hough table. Thus the remaining three transformation parameters still must perform a good deal of work to eliminate the false matches. Even with the full 6 parameters, false matches sometimes remain [Thompson and Mundy 87]. Finally, the amount of search required is very large, as about 10 million buckets must be considered in order to find the buckets with peaks.

In order to get a more complete picture of the utility of the generalized Hough transform for transformation clustering, Table 18 shows how large the peak size, l, corresponding to a correct match must be in order to limit the probability of a random peak of at least that size, $p_{\geq l}$. The values are shown for various levels of $\lambda = \frac{r}{n}$, the ratio of number of table entries to number of buckets are shown, and various probabilities, $p_{\geq l}$. Recall that in order for the expected number of false matches to be less than 1, the probability should be less than $\frac{1}{n}$. Thus for a Hough table with 10,000 entries the corresponding column would be 10^{-4}, and for a million entries it is the 10^{-6} column.

$p_{\geq l} =$	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
$\lambda = .25$	2	3	4	5	6	7
.5	3	4	5	6	7	8
1	4	5	6	8	9	10
2	6	8	9	10	12	13
4	9	11	13	15	17	18
8	15	18	20	23	25	27
16	26	30	33	36	38	41
32	46	51	55	59	62	65

Table 18. Peak size, l, for different values of $\lambda = \frac{r}{n}$, and different probabilities, $P_{\geq l}$, of peaks at least as large as l occurring at random.

5. Summary

We have formally considered several aspects of the generalized Hough transform as a method for recognizing objects from noisy data in complex cluttered environments. We have analyzed both the redundancy of the bucketing operation, and the likelihood that random clusters of transformations will be as large as those resulting from a correct match. The major results of this analysis are as follows:

1. We have shown analytically that the range of transformations specified by a given pairing of model and image features, can be quite large. This is particularly true in the case of extended features, which can be partially occluded in the scene, and in the presence of significant amounts of sensor uncertainty.

2. We have shown analytically, and through representative examples, that the number of Hough buckets specified by such a range of transformations can also be quite large. The fraction of the total number of buckets that are specified by a single data-model pairing increases with increasing sensor uncertainty, with a reduction in the total number of buckets (i.e. increasing coarseness of the Hough space), with increasing occlusion, when projections onto subspaces of the full parameter space are used, and when scale is allowed to vary.

3. We have shown, using an occupancy model, that the number of model-image pairings likely to fall into the same Hough bucket at random, can be quite high. As a consequence the clusters that occur at random are often likely to be larger than those that correspond to a correct solution. This may force a recognition system to examine large portions of the Hough space, in order to verify a correct interpretation from a spurious collection of parameter vectors. This problem is exacerbated as the redundancy factor increases, and hence is affected by changes in sensor uncertainty, Hough tessellation and scene complexity, as above.

Our conclusion is that while the generalized Hough transform technique is useful for some classes of recognition tasks, it does not scale well, and is poorly suited to recognition in complex environments. For example, our analysis suggests that the Hough transform should be adequate for the recognition of objects with limited occlusion and moderate sensor uncertainty, using isolated points such as vertices as the matching features. This is supported by the empirical evidence of several researchers in the field (e.g., [Silberberg et. al. 84] [Linainmaa et al. 85]). At the same time, however, the analysis suggests that the method will scale poorly, when applied to complex, cluttered scenes, or when using extended features such as edges (which are subject to partial occlusion).

It may seem somewhat surprising that the expected performance of the generalized Hough transform is so poor for complex images. Recall, however, that the operation was originally used to separate outliers from good data. Its first use in recognition was for relatively simple tasks, where the data corresponding to the correct solution is a fairly large fraction of all the data. In contrast, for recognition in complex scenes the good data is a small fraction of the incorrect data, or "outliers". It just turns out that the method does not scale very well to tasks where the amount of correct data is relatively small compared to the amount of incorrect data.

References

V. S. Alagar and L. H. Thiel, 1981. Algorithms for detecting M-dimensional objects in N-dimensional spaces. *IEEE Trans. on Pattern Analysis and Machine Intelligence,* **3**(3):245–256.

D. H. Ballard, 1981. Generalizing the Hough transform to detect arbitrary shapes. *Pattern Recognition* **13**:111.

P. J. Besl and R. C. Jain, 1985. Three-dimensional object recognition. *ACM Computing Surveys,* **17**(1):75–154.

C. M. Brown, 1983. Inherent bias and noise in the Hough transform. *IEEE Trans. on Pattern Analysis and Machine Intelligence,* **5**(5):493–505.

R. T. Chin and C. R. Dyer, 1986. Model-based recognition in robot vision. *ACM Computing Surveys,* **18**(1):67–108.

D. T. Clemens, 1986. *The Recognition of Two-Dimensional Modeled Objects in Images,* M.S. Thesis, Dept. of Electrical Engineering and Computer Science, Massachusetts Institute of Technology.

M. Cohen and G. T. Toussaint, 1977. On the detection of structures in noisy pictures. *Pattern Recognition* **9**:95–98.

L. S. Davis, 1982. Hierarchical generalized Hough transforms and line-segment based generalized Hough transforms. *Pattern Recognition* **15**:277.

W. Feller, 1968. *An Introduction to Probability Theory and Its Applications*, Wiley, New York.

Y. Lamdan, J. T. Schwartz and H. J. Wolfson, 1987. On recognition of 3-d objects from 2-d images, New York University, Courant Institute Robotics Report, No. 122.

S. Linainmaa, D. Harwood, and L. S. Davis, 1985. Pose determination of a three-dimensional object using triangle pairs, CAR-TR-143, Center for Automation Research, University of Maryland.

H. Maitre, 1986. Contribution to the prediction of performances of the Hough transform. *IEEE Trans. on Pattern Analysis and Machine Intelligence,* **8**(5):669–674.

S. D. Shapiro, 1975. Transformations for the computer detection of curves in noisy pictures. *Computer Graphics and Image Processing,* 4:328–338.

S. D. Shapiro and A. Iannino, 1979. Geometric constructions for predicting Hough transform performance. *IEEE Trans. on Pattern Analysis and Machine Intelligence,* **1**(3):310–317.

T. M. Silberberg, L. S. Davis, and D. A. Harwood, 1984. An iterative Hough procedure for three-dimensional object recognition. *Pattern Recognition* **17**(6):612–629.

T. M. Silberberg, D. A. Harwood, and L. S. Davis, 1986. Object recognition using oriented model points. *Computer Vision, Graphics, and Image Processing* **35**:47–71.

D. W. Thompson and J. L. Mundy, 1987. Three-dimensional model matching from an unconstrained viewpoint. *Proceedings of the International Conference on Robotics and Automation*, IEEE Computer Society Press, pp. 208–220.

J. L. Turney, T. N. Mudge and R. A. Volz, 1985. Recognizing partially occluded parts, *IEEE Trans. Pattern Anal. and Machine Intel.* 7(4):410–421.

T. M. van Veen and F. C. A. Groen, 1981. Discretization errors in the Hough transform. *Pattern Recognition* 14:137–145.

Appendix

Analysis of the basic case

In this section, we fully derive the relationship defining the redundancy of the Hough transform for two dimensional edge segments.

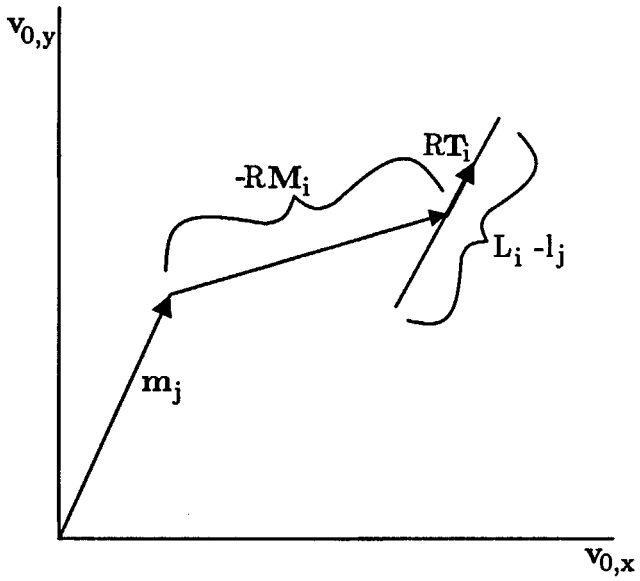

Figure 5. Range of feasible translations, for fixed θ and with no position error. The line in the direction of $R\mathbf{T}_i$ denotes the set of feasible translations for a given value of θ.

Suppose we are considering the matching of a data edge with a model edge. Consider the situation shown in Figure 5. This shows the set of consistent translations, for a given value of θ, say θ_c, where we ignore for now the effect of error ϵ_p. That is, for a given rotation, equation (1) defines a set of translations, which are illustrated in the figure. Now, as θ varies, this line will vary, in particular, it will rotate about the center defined by $\mathbf{m}_j$, with a radius of $\|\mathbf{M}_j\|$. We want to determine the union of the projection of each such line into the x–y plane. The situation is shown in Figure 6.

To find the area of this region, we use the following simple trick. Consider the lower hashed region shown in Figure 7. If we translate and rotate this region to the upper hashed region shown in the figure, then we see that the area of the remaining region is simply given by

$$\int_{\theta=0}^{h_\theta} \int_{\rho=S_\ell}^{S_h} \rho \, d\rho \, d\theta = \frac{{S_h}^2 - {S_\ell}^2}{2} h_\theta.$$

To derive the limits S_ℓ and S_h, we can use the parameters from the known edges. Consider Figure 8. Here, M_J denotes the size of the vector $\|\mathbf{M}_J\|$, and ϕ is the angle

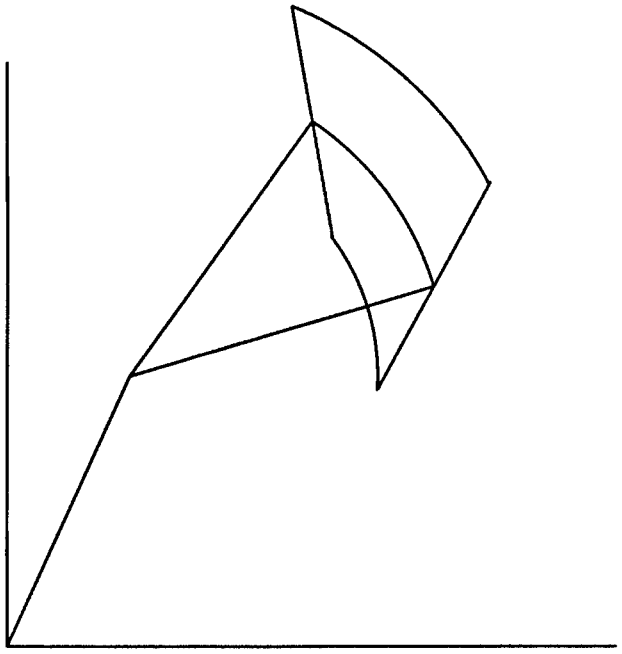

Figure 6. Rotation of the line of Figure 5 through h_θ radians.

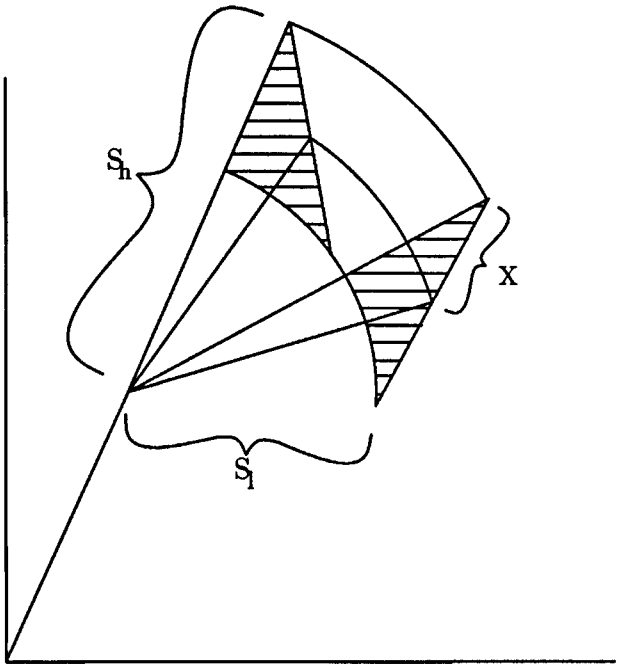

Figure 7. Total area of the swept region.

made between $\mathbf{M}_J$ and $\hat{\mathbf{T}}_J$. The distance X is simply given by $X = (L_J - \ell_j)/2$.

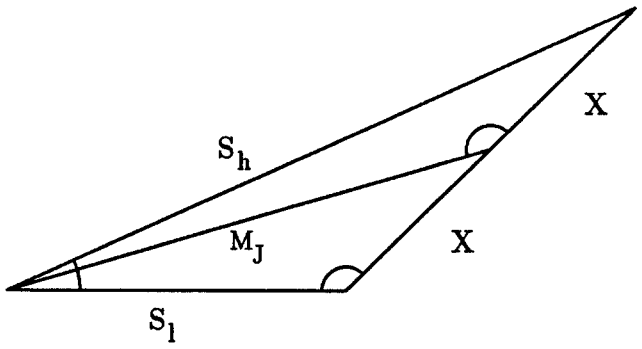

Figure 8.

Using the law of cosines, we have

$$S_\ell{}^2 = M_J^2 + X^2 - 2M_J X\cos(\pi - \phi)$$
$$S_h{}^2 = M_J^2 + X^2 - 2M_J X\cos(\phi).$$

Thus, the entire area covered is

$$A = 2M_J h_\theta X \sin\left(\phi - \frac{\pi}{2}\right).$$

Note that by symmetry, we can assume that $\phi \in [\pi/2, \pi]$, as the other cases are similar.

This analysis, however, ignored the effects of the sensing error. In particular, we know that the translation can be determined only to within a ball of radius ϵ_p. Thus, the full area is swept out by first sweeping this ball along the line of feasible translations, and then sweeping that entire region through the angle h_θ. This is equivalent to expanding the region swept out by rotating the line over h_θ to include any point within a distance ϵ_p of the boundary of this region. The additional area is shown in Figure 9.

The largest circular piece (denote (1) in the figure) has an area given by

$$\int_{\theta=0}^{h_\theta}\int_{\rho=S_h}^{S_h+\epsilon_p} \rho d\rho d\theta = h_\theta \frac{(S_h+\epsilon_p)^2 - S_h{}^2}{2}.$$

Similarly, the smaller circular piece (denote (2) in the figure) has area

$$h_\theta \frac{S_\ell{}^2 - (S_\ell - \epsilon_p)^2}{2}.$$

The two rectangular pieces (denoted (3) and (4) in the figure) have area

$$4X\epsilon_p.$$

Finally, the four joining segments have a total angular extent of 2π so that they contribute an area of

$$\pi\epsilon_p^2.$$

Combining these areas with the original area, we find that the area covered by the entire region is

$$A(h_\theta, \epsilon_p, M_J, L_J, \ell_j, \phi) = 2M_J h_\theta X \sin\left(\phi - \frac{\pi}{2}\right) + \pi\epsilon_p^2 + 4\epsilon_p X + \epsilon_p h_\theta [S_h + S_\ell].$$

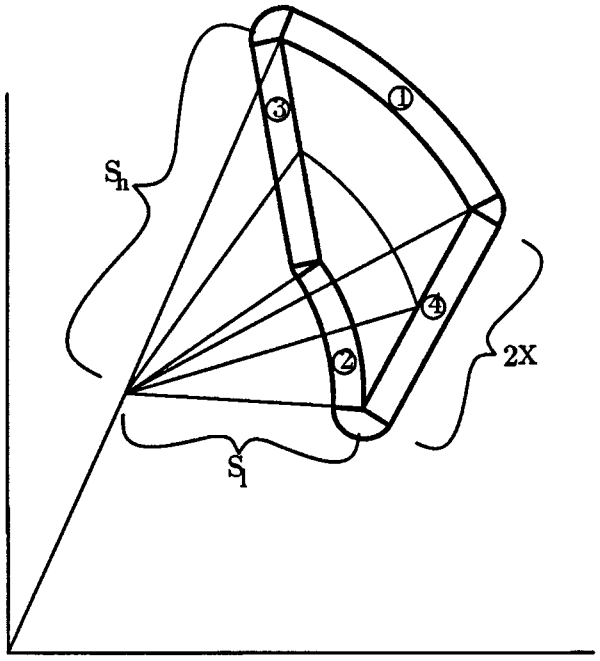

Figure 9. Additional region of feasible translations due to sensing error.

Now, we need to find a lower bound for the number of buckets that are intersected by such an area. The simplest lower bound, which is not a tight one, is given by assuming that the area is square, and can be tightly packed into the x–y portion of the Hough buckets. This may badly underestimate the number of buckets intersected by a volume, but it provides a convenient starting place. If the region is a tightly packed square, then the minimum number of buckets is given by

$$\left\lceil \frac{A(h_\theta, \epsilon_p, M_J, L_J, \ell_j, \phi)}{h_t^2} \right\rceil .$$

Now, this region corresponds to the number of buckets intersected, as the rotation component varies over the dimension of a single bucket. Thus, the redundancy factor for pose clustering, b, i.e. the number of buckets into which a single data-model pairing casts a vote is bounded below by

$$b \geq \left\lceil \frac{2\epsilon_a}{h_\theta} \right\rceil \left\lceil \frac{A(h_\theta, \epsilon_p, M_J, L_J, \ell_j, \phi)}{h_t^2} \right\rceil$$

where the bound on angular error is given by

$$\epsilon_a = \tan^{-1}\left(\frac{2\epsilon_p}{\sqrt{\ell^2 - 4\epsilon_p^2}} \right)$$

and where the area is given by

$$A(h_\theta, \epsilon_p, M_J, L_J, \ell_j, \phi) = M_J h_\theta (L_J - \ell_j) \sin\left(\phi - \frac{\pi}{2}\right) + \pi\epsilon_p^2 + 2\epsilon_p(L_J - \ell_j) + \epsilon_p h_\theta [S_h + S_\ell].$$

The measurements in which we are interested depend on the relationship between dimensions of the object and the tesselation of the Hough space. We can simplify our expressions, by using relative measurements. In particular, we let

$$\epsilon_p^* = \frac{\epsilon_p}{h_t} \qquad L^* = \frac{L}{h_t} \qquad \ell^* = \frac{\ell}{h_t} \qquad M^* = \frac{M}{h_t}$$

so that the redundancy of the Hough space is

$$b \geq \left\lceil \frac{2\epsilon_a}{h_\theta} \right\rceil \left\lceil A^*(h_\theta, \epsilon_p^*, M_J^*, L_J^*, \ell_j^*, \phi) \right\rceil$$

where

$$\epsilon_a = \tan^{-1}\left(\frac{2\epsilon_p^*}{\sqrt{(\ell^*)^2 - 4\left(\epsilon_p^*\right)^2}} \right)$$

$$A^*(h_\theta, \epsilon_p^*, M_J^*, L_J^*, \ell_j^*, \phi) = M_J^* h_\theta (L_J^* - \ell_j^*) \sin\left(\phi - \frac{\pi}{2}\right) + \pi\left(\epsilon_p^*\right)^2 + 2\epsilon_p^*(L_J^* - \ell_j^*) + \epsilon_p^* h_\theta [S_h^* + S_\ell^*]$$

$$S_h^* = \sqrt{(M_J^*)^2 + \frac{1}{4}\left(L_J^* - \ell_j^*\right)^2 - M_J^*\left(L_J^* - \ell_j^*\right)\cos\phi}$$

$$S_\ell^* = \sqrt{(M_J^*)^2 + \frac{1}{4}\left(L_J^* - \ell_j^*\right)^2 + M_J^*\left(L_J^* - \ell_j^*\right)\cos\phi}.$$

This gives careful bounds on the redundancy factor. We can get more useful bounds by considering the following case. We will assume that angle ϕ between $\mathbf{M}_J$ and $\hat{\mathbf{T}}_J$ is uniformly distributed over the range

$$\left[\frac{\pi}{2}, \pi\right].$$

This allows us to estimate the expected value of the first term for A^*. Finding the expected value for S_h and S_ℓ involves elliptic integrals of the second kind, so we underestimate the area by finding the minimum value for $S_h + S_\ell$, as ϕ varies over its range. A straightforward application of the calculus leads to:

$$S_h^* + S_\ell^* \geq 2M^*.$$

We will also assume that $L_J = L$ for all model edges, that $M_J = M$ for all model edges, and that the data edges are of equal length,

$$\ell_j = \beta L$$

for some parameter β, $\frac{2\epsilon_p}{L} \leq \beta \leq 1$.

Under these conditions, the expected area is at least

$$A^*(h_\theta, \epsilon_p^*, M^*, L^*, \beta) \geq 2M^*(1-\beta)L^*\frac{h_\theta}{\pi} + \pi\left(\epsilon_p^*\right)^2 + 2\epsilon_p^* L^*(1-\beta) + 2\epsilon_p^* h_\theta M^* \quad (5a)$$

and the expected redundancy is at least

$$b \geq \left\lceil \frac{2\epsilon_a}{h_\theta} \right\rceil \left\lceil A^*(h_\theta, \epsilon_p^*, M^*, L^*, \beta) \right\rceil \quad (5b)$$

where the bound on angular error is given by

$$\epsilon_a = \tan^{-1}\left(\frac{2\epsilon_p^*}{\sqrt{(\ell^*)^2 - 4(\epsilon_p^*)^2}} \right). \quad (5c)$$

The expressions in equation (5) give a lower bound on the expected number of buckets intersected by a data-model pairing. This lower bound is not tight, as in deriving it we have assumed that the area of consistency in the x–y plane can be tightly packed into the tesselations of the Hough space. A better lower bound on the expected number of buckets can be obtained by accounting for the fact that the area of consistency may only partially intersect buckets along its border. An example is shown in Figure 10, in which the swept region has an area that is roughly equivalent to 6 buckets in size, but which actually intersects 14 different buckets.

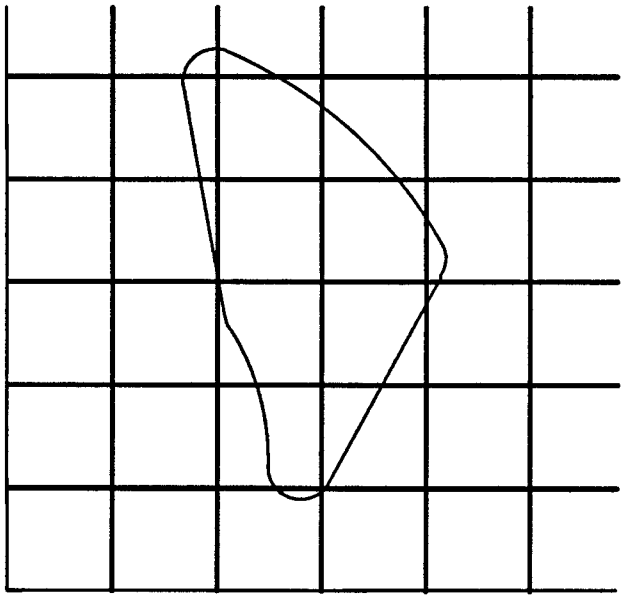

Figure 10. The number of buckets intersected may be larger than the ratio of the area of the region to the area of a bucket.

A simple means of accounting for this effect is to observe that on average, a bucket on the border of the swept region will be only half occupied. As well, the perimeter of the swept region can be easily shown to be

$$P = (S_h + S_\ell)\, h_\theta + 2\,(L_J - \ell_j) + 2\pi\epsilon_p$$

which is bounded, by our earlier analysis, by

$$P \geq 2M_J h_\theta + 2\,(L_J - \ell_j) + 2\pi\epsilon_p.$$

The minimum number of buckets intersected by this perimeter is

$$\frac{P}{\sqrt{2}h_t}.$$

If we normalize with respect to the bucket size, we have

$$P^* \geq 2M_J^* h_\theta + 2\left(L_J^* - \ell_j^*\right) + 2\pi\epsilon_p^*.$$

Since, on average, border buckets are half occupied, in place of A^*, we can now use

$$A^* + \frac{P^*}{2\sqrt{2}}$$

If we let

$$A_+^*(h_\theta, \epsilon_p^*, M^*, L^*, \beta) = A^* + \frac{P^*}{2\sqrt{2}}$$
$$\geq 2M^*(1-\beta)L^*\frac{h_\theta}{\pi} + \pi\left(\epsilon_p^*\right)^2 + 2\epsilon_p^* L^*(1-\beta) + 2\epsilon_p^* h_\theta M^*$$
$$+ \frac{1}{2\sqrt{2}}\left(2M^* h_\theta + 2(1-\beta)L^* + 2\pi\epsilon_p^*\right) \tag{3a}$$

then the expected redundancy is at least

$$b \geq \left\lceil \frac{2\epsilon_a}{h_\theta} \right\rceil \lceil A_+^* \rceil \tag{3b}$$

where the bound on angular error is given by

$$\epsilon_a = \tan^{-1}\left(\frac{2\epsilon_p^*}{\sqrt{(\ell^*)^2 - 4(\epsilon_p^*)^2}}\right). \tag{3c}$$

Analysis of the scaled case

Similar to the case of rigid objects, we need to formally derive the redundancy of the Hough transform for objects that can freely scale.

To determine the redundancy factor for parameter hashing in the case of scale, we again want to determine the number of buckets consistent with a data-model pairing for a single slice of the x–y components of the transform space. Note that in this case, the transform space $\mathcal{T}$ is four dimensional, with an extra axis for the scale factor. Projecting the volume obtained as θ varies over the bounds of a single bucket gives us the volume shown in Figure 6, where now the borders are functions of the scale factor. If we now look at the projection of the volume as k is varied, we will get the region obtained by varying the region in Figure 6 over the range of values of k. This new region is shown in Figure 3. We need to determine the area spanned by this region. The heavy lines in Figure 3 break the total area into three portions. The previous analysis implies that the large portion has an area

$$2h_\theta M_J'(k_h) X_{ij}'(k_h) \sin\left(\phi - \frac{\pi}{2}\right)$$

where

$$M_J'(k) = kM_J$$
$$X_{ij}'(k) = \frac{kL_J - \ell_j}{2}$$

and where k varies from k_ℓ to k_h and M_J is the midpoint distance of the model face without any scaling.

The circular segment of the area in Figure 3 has an area given by

$$\int_{\theta=0}^{h_\theta}\int_{\rho=S_\ell(k_\ell)}^{S_\ell(k_h)} \rho d\rho d\theta = h_\theta \frac{S_\ell(k_h)^2 - S_\ell(k_\ell)^2}{2}$$

where

$$S_\ell(k)^2 = M'_J(k)^2 + X'_{ij}(k)^2 + 2M'_J(k)X'_{ij}(k)\cos\phi$$
$$S_h(k)^2 = M'_J(k)^2 + X'_{ij}(k)^2 - 2M'_J(k)X'_{ij}(k)\cos\phi.$$

To get the area of the slice of the triangular portion, we use the law of sines to derive

$$\left[X'_{ij}(k_h)M'_J(k_h) - X'_{ij}(k_\ell)M'_J(k_\ell)\right]\sin\phi.$$

Hence, the area of consistent translations, ignoring the error ϵ_p is given by

$$\begin{aligned} A_s =& h_\theta k_h M_J(k_h L_J - \ell_j)\sin\left(\phi - \frac{\pi}{2}\right) \\ &+ h_\theta(k_h - k_\ell)\left[\frac{k_h + k_\ell}{2}\left(M_J{}^2 + \frac{L_J{}^2}{4}\right) - \frac{L_J\ell_j}{2}\right] \\ &- h_\theta\frac{M_J}{2}\cos\phi\left[(k_h^2 + k_\ell^2)L_j - (k_h + k_\ell)\ell_j\right] \\ &+ M_J(k_h - k_\ell)\left[\frac{k_h + k_\ell}{2}L_J - \frac{\ell_j}{2}\right]\sin\phi. \end{aligned}$$

We must also account for error in measuring the position. As in the previous case, the addition in this case is found by expanding the area in Figure 3 by a distance ϵ_p, as shown in Figure 11.

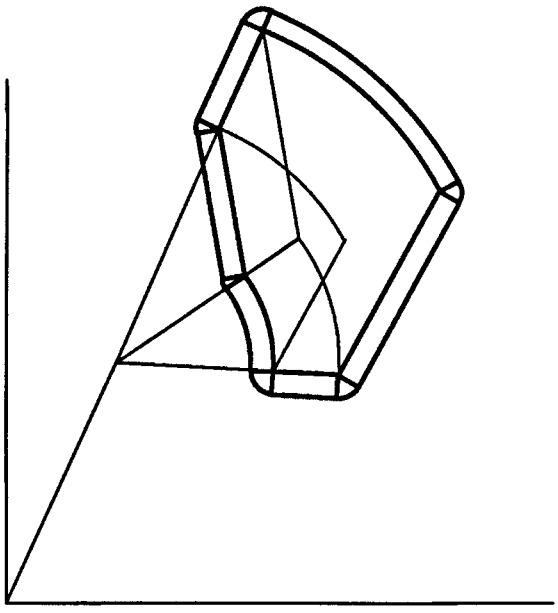

Figure 11. Region of translation space consistent with scale variation and angle variation.

Using techniques similar to those employed in the case of no scaling, we find that the additional area, due to sensing error is given by

$$\begin{aligned} &h_\theta\epsilon_p\left[S_h(k_h) + S_\ell(k_\ell)\right] + \epsilon_p\left[(k_h + k_\ell)L_J - 2\ell_j\right] \\ &+ \epsilon_p\left[S_h(k_h) - S_h(k_\ell) + S_\ell(k_h) - S_\ell(k_\ell)\right] + \pi\epsilon_p^2. \end{aligned}$$

Combining these two results yields an area of

$$\begin{aligned}
A_s =& h_\theta k_h M_J (k_h L_J - \ell_j) \sin\left(\phi - \frac{\pi}{2}\right) \\
&+ h_\theta (k_h - k_\ell) \left[\frac{k_h + k_\ell}{2}\left(M_J{}^2 + \frac{L_J{}^2}{4}\right) - \frac{L_J \ell_j}{2}\right] \\
&- h_\theta \frac{M_J}{2} \cos\phi \left[(k_h^2 + k_\ell^2) L_j - (k_h + k_\ell)\ell_j\right] \\
&+ M_J (k_h - k_\ell) \left[\frac{k_h + k_\ell}{2} L_J - \frac{\ell_j}{2}\right] \sin\phi \\
&+ h_\theta \epsilon_p [S_h(k_h) + S_\ell(k_\ell)] + \epsilon_p [(k_h + k_\ell) L_J - 2\ell_j] \\
&+ \epsilon_p [S_h(k_h) - S_h(k_\ell) + S_\ell(k_h) - S_\ell(k_\ell)] + \pi \epsilon_p^2.
\end{aligned}$$

Similar to the case of no scale, we can bound this below by finding the minimum value taken on by the S_h and S_ℓ terms, yielding

$$\begin{aligned}
A_s \geq& h_\theta k_h M_J (k_h L_J - \ell_j) \sin\left(\phi - \frac{\pi}{2}\right) \\
&+ h_\theta (k_h - k_\ell) \left[\frac{k_h + k_\ell}{2}\left(M_J{}^2 + \frac{L_J{}^2}{4}\right) - \frac{L_J \ell_j}{2}\right] \\
&- h_\theta \frac{M_J}{2} \cos\phi \left[(k_h^2 + k_\ell^2) L_j - (k_h + k_\ell)\ell_j\right] \\
&+ M_J (k_h - k_\ell) \left[\frac{k_h + k_\ell}{2} L_J - \frac{\ell_j}{2}\right] \sin\phi \\
&+ h_\theta \epsilon_p \left[(k_h + k_\ell) M_J - \frac{k_h - k_\ell}{2} L_J\right] + \epsilon_p [(k_h + k_\ell) L_J - 2\ell_j] \\
&+ 2\epsilon_p (k_h - k_\ell) M_J + \pi \epsilon_p^2.
\end{aligned}$$

As before, we can take the expected value of this expression as ϕ varies uniformly over the range $[\pi/2, \pi]$. This region corresponds to the number of buckets intersected as the rotation component varies over the range of a single bucket, and as the sacle factor varies over the range of a single bucket. Note that in this case, the area of the translation component of the Hough space that is consistent with an assignment is actually a function of the scale factor, rather than just a function of the size of the Hough buckets and the properties of the object and the sensing errors. We can

rewrite this equation in terms of the range of variation in scale, Δk:

$$\begin{aligned}
A_s(\Delta k, k_h) =& h_\theta k_h M_J (k_h L_J - \ell_j) \frac{2}{\pi} \\
&+ h_\theta \Delta k \left[\left(k_h - \frac{\Delta k}{2} \right) \left(M_J^2 + \frac{L_J^2}{4} \right) - \frac{L_J \ell_j}{2} \right] \\
&+ \frac{h_\theta}{\pi} M_J \left[(2k_h^2 + 2k_h \Delta k + \Delta k^2) L_j - (2k_h + \Delta k) \ell_j \right] \\
&+ M_J \Delta k \left[\left(k_h - \frac{\Delta k}{2} \right) L_J - \frac{\ell_j}{2} \right] \frac{2}{\pi} \\
&+ h_\theta \epsilon_p \left[(2k_h - \Delta k) M_J - \frac{\Delta k}{2} L_J \right] + 2\epsilon_p \left(2 \left(k_h - \frac{\Delta k}{2} \right) L_J - 2\ell_j \right) \\
&+ 2\epsilon_p \Delta k M_J + \pi \epsilon_p^2.
\end{aligned}$$

As in the previous case, we can normalize the measurements relative to the dimensions of the Hough spacing, h_t, so that the area is given by

$$\begin{aligned}
A_s^*(\Delta k, k_h) =& h_\theta k_h M_J^* (k_h L_J^* - \ell_j^*) \frac{2}{\pi} \\
&+ h_\theta \Delta k \left[\left(k_h - \frac{\Delta k}{2} \right) \left((M_J^*)^2 + \frac{(L_J^*)^2}{4} \right) - \frac{L_J^* \ell_j^*}{2} \right] \\
&+ \frac{h_\theta}{\pi} M_J^* \left[(2k_h^2 + 2k_h \Delta k + \Delta k^2) L_j^* - (2k_h + \Delta k) \ell_j^* \right] \\
&+ M_J^* \Delta k \left[\left(k_h - \frac{\Delta k}{2} \right) L_J^* - \frac{\ell_j^*}{2} \right] \frac{2}{\pi} \\
&+ h_\theta \epsilon_p^* \left[(2k_h - \Delta k) M_J^* - \frac{\Delta k}{2} L_J^* \right] + 2\epsilon_p^* \left(2 \left(k_h - \frac{\Delta k}{2} \right) L_J^* - 2\ell_j^* \right) \\
&+ 2\epsilon_p^* \Delta k M_J^* + \pi (\epsilon_p^*)^2.
\end{aligned}$$

Suppose we define the full range of possible scale factors to be $[1, k_{max}]$, so that the model is defined as the smallest possible instance of an object. Then to count the redundancy factor in this case, we must sum the number of buckets obtained over all possible scale factors. If the spacing of the Hough buckets in the scale dimension is h_k, then this sum is given by:

$$b_s = \left\lceil \frac{2\epsilon_a}{h_\theta} \right\rceil \left\lceil A_s^*(i_s - \max\{1, \frac{\ell^*}{L^* h_k}\}, i_s h_k) \right\rceil + \sum_{i=i_s}^{\frac{k_{max}}{h_k}} \left\lceil \frac{2\epsilon_a}{h_\theta} \right\rceil \lceil A_s^*(h_k, i h_k) \rceil$$

where

$$i_s = \left\lceil \frac{\max(1, \frac{\ell^*}{L})}{h_k} \right\rceil$$

is the starting point for the scale summation, and where the first term in the expression captures any partial inclusion of a bucket.

We have assumed that k_{max} is some integer multiple of h_k. As before, the

bound on angular error is given by

$$\epsilon_a = \tan^{-1}\left(\frac{2\epsilon_p^*}{\sqrt{(\ell^*)^2 - 4\left(\epsilon_p^*\right)^2}}\right).$$

Similar to the non-scaled case, we can obtain tighter bounds by considering the buckets on the edge of the region, which are likely to be only partially intersected by the region. The perimeter of Figure 9 can be shown to equal

$$\begin{aligned} P_s =& h_\theta\left(S_\ell(k_\ell) - \epsilon_p\right) + k_\ell L_J - \ell_j + S_h(k_h) - S_h(k_\ell) \\ & h_\theta\left(S_h(k_h) + \epsilon_p\right) + k_h L_J - \ell_j + S_\ell(k_h) - S_\ell(k_l) + 2\pi\epsilon_p \end{aligned}$$

and this is bounded by

$$P_s \geq h_\theta\left((k_h + k_\ell)M_J - \frac{k_h - k_\ell}{2}L_J\right) + (k_h + k_\ell)L_J - 2\ell_j + 2(k_h - k_\ell)M_J + 2\pi\epsilon_p.$$

If we normalize with respect to bucket size, we get

$$P_s^* \geq h_\theta\left((k_h + k_\ell)M_J^* - \frac{k_h - k_\ell}{2}L_J^*\right) + (k_h + k_\ell)L_J^* - 2\ell_j^* + 2(k_h - k_\ell)M_J^* + 2\pi\epsilon_p^*.$$

Hence, a better bound on the expected redundancy is given by

$$b_s = \left\lceil\frac{2\epsilon_a}{h_\theta}\right\rceil\left\lceil A_{s,+}^*(i_s - \max\{1, \frac{\ell^*}{L^* h_k}\}, i_s h_k)\right\rceil + \sum_{i=i_s}^{\frac{k_{max}}{h_k}}\left\lceil\frac{2\epsilon_a}{h_\theta}\right\rceil\left\lceil A_{s,+}^*(h_k, ih_k)\right\rceil$$

where

$$i_s = \left\lceil\frac{\max(1, \frac{\ell^*}{L})}{h_k}\right\rceil$$

and where

$$A_{s,+}^* = A_s^* + \frac{P_s^*}{2\sqrt{2}}.$$

This blank page was inserted to preserve pagination.

CS-TR Scanning Project
Document Control Form

Date : 4/27/95

Report # AIM-1044

Each of the following should be identified by a checkmark:
Originating Department:

☒ Artificial Intellegence Laboratory (AI)
☐ Laboratory for Computer Science (LCS)

Document Type:

☐ Technical Report (TR) ☒ Technical Memo (TM)
☐ Other:_____

Document Information

Number of pages: 41 (47-IMAGES)
Not to include DOD forms, printer instructions, etc... original pages only.

Originals are:
☒ Single-sided or
☐ Double-sided

Intended to be printed as :
☐ Single-sided or
☒ Double-sided

Print type:
☐ Typewriter ☐ Offset Press ☐ Laser Print
☐ InkJet Printer ☐ Unknown ☐ Other:_____

Check each if included with document:

☒ DOD Form (2) ☐ Funding Agent Form ☐ Cover Page
☐ Spine ☐ Printers Notes ☐ Photo negatives
☐ Other: _____

Page Data:

Blank Pages (by page number):_____

Photographs/Tonal Material (by page number):_____

Other (note description/page number):

	Description :	Page Number:
(A)	IMAGE MAP (1) UN#'ED TITLE PAGE	
	(2-41) PAGES #'ED 1-40	
	(42-44) SCANCONTROL, DOD (2)	
	(45-47) TRGTJ (3)	
(B)	MARK IN UPPER RIGHT CORNER OF EACH PAGE	

Scanning Agent Signoff:

Date Received: 4/27/95 Date Scanned: 5/2/95 Date Returned: 5/4/95

Scanning Agent Signature: Michael N. Cook

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AIM 1044	2. GOVT ACCESSION NO. AD-A202372	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) On the Sensitivity of the Hough Transform for Object Recognition		5. TYPE OF REPORT & PERIOD COVERED memorandum 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) W. Eric L. Grimson and Daniel P. Huttenlocher		8. CONTRACT OR GRANT NUMBER(s) N00014-86-K-0685 DACA76-85-C-0010 N00014-85-K-0124
9. PERFORMING ORGANIZATION NAME AND ADDRESS Artificial Intelligence Laboratory 545 Technology Square Cambridge, MA 02139		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Advanced Research Projects Agency 1400 Wilson Blvd. Arlington, VA 22209		12. REPORT DATE May 1988 13. NUMBER OF PAGES 40
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Office of Naval Research Information Systems Arlington, VA 22217		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution is unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

None

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Hough transform
object recognition

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Object recognition from sensory data involves, in part, determining the pose of a model with respect to a scene. A common method for finding an object's pose is the generalized Hough transform, which accumulates evidence for possible coordinate transformations in a parameter space whose axes are the quantized transformation parameters. Large clusters of similar transformations in that space are taken as evidence of a correct match. In this article, we provide a theoretical analysis of the behavior of such methods. We derive bounds on the set of transformations consistent with each pairing of data and model features, in

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: AIM 1044	2. GOVT ACCESSION NO: AD-A202 342
4. TITLE (and Subtitle): [illegible]	3. RECIPIENT'S CATALOG NUMBER:
7. AUTHOR(s): W. Eric L. Grimson and Daniel P. Huttenlocher	8. CONTRACT OR GRANT NUMBER(s): N00014-86-K-0685 DACA76-85-C-0010 N00014-85-K-0124
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Artificial Intelligence Laboratory 545 Technology Square Cambridge, MA 02139	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:
11. CONTROLLING OFFICE NAME AND ADDRESS: Advanced Research Projects Agency 1400 Wilson Blvd. Arlington, VA 22209	12. REPORT DATE: May 1988
	13. NUMBER OF PAGES: 40
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office): Office of Naval Research Information Systems Arlington, VA 22217	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECL ASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report)

Distribution is unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

None

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Hough transform
object recognition

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Object recognition from sensory data involves, in part, determining the pose of a model with respect to a scene. A common method for finding an object's pose is the generalized Hough transform, which accumulates evidence for possible coordinate transformations in a parameter space where the quantized transformation parameters. Large clusters of similar transformations in that space are taken as evidence of a correct match. In this article, we provide a theoretical analysis of the behavior of such methods. We derive bounds on the sort of transformations consistent with each pairing of data and model features. In

Block 20, cont.

the presence of noise and occlusion in the image. We also provide bounds on the likelihood of false peaks in the parameter space, as a function of noise, occlusion, and tessellation effects. We argue that [illegible] with the complexity on recognition tasks in a similar proposition, as the probability of false [illegible] can be very high.

DD FORM 1473 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

UNCLASSIFIED

Scanning Agent Identification Target

Scanning of this document was supported in part by the **Corporation for National Research Initiatives,** using funds from the **Advanced Research Projects Agency** of the **United states Government** under Grant: **MDA972-92-J1029.**

The scanning agent for this project was the **Document Services** department of the **M.I.T Libraries.** Technical support for this project was also provided by the **M.I.T. Laboratory for Computer Sciences.**