Probabilistic Indexing: Recognizing 3D Objects from 2D Images Using the Probabilistic Peaking Effect

Clark F. Olson
Computer Science Division
University of California at Berkeley
Berkeley, CA 94720
clarko@robotics.berkeley.edu

Abstract

Recent papers [Lamdan *et al.*, 1988, Clemens and Jacobs, 1991] have shown that indexing is a promising approach to fast model-based object recognition because it allows most of the possible matches between image point groups and model point groups to be quickly eliminated from consideration. Current indexing systems for the problem of recognizing three-dimensional objects from single two-dimensional images require groups of four points to generate a key into the table of model groups and each model group must be represented over an infinite subspace of a multi-dimensional table [Clemens and Jacobs, 1991, Jacobs, 1992]. We present a system that is capable of indexing using groups of three points by taking advantage of the probabilistic peaking effect [Ben-Arie, 1990]. Each model group need only be represented at one point in the index table. To be able to index using groups of three points, we must allow false negatives for point group matches. If there are n model points present in the image, there are $O(n^3)$ groups of three correct model points, so we can withstand false negatives by examining information from multiple groups. Since we are able to index on smaller groups of points, indexing can be used with an additional set of algorithms with lower computational complexity. This system can utilize larger point groups to increase accuracy in discriminating between correct and incorrect matches. Results are given on real and synthetic data.

1 Introduction

In the model-based object recognition paradigm, a catalog of object models is used to recognize the objects in images. This paper discusses techniques in this paradigm for the case of recognizing three-dimensional objects represented by feature points from a single image of two-dimensional data. Some methods that have been used to attack the object recognition problem are:

- Search Methods: These search the space of matches between image features and model features to determine mutually feasible matches [Faugeras and Hebert, 1986, Ben-Arie and Meiri, 1987, Grimson, 1990]. These techniques generally require a run time that is exponential in the number of image feature.

- Alignment Methods: These examine matches of small groups of model and image features to determine the transformations that align them. Verification is then performed to determine if the transformations is correct [Ayache and Faugeras, 1986, Lowe, 1987, Huttenlocher and Ullman, 1990].

- Pose Methods: These examine transformation space to determine transformations that bring a large number of image features and model features into correspondence [Ballard, 1981, Thompson and Mundy, 1987, Cass, 1990, Grimson *et al.*, 1992].

A tool that can be useful to all of the recognition methods described above is indexing. Indexing determines which groups of model points could have projected to specific groups of image points, eliminating the need to consider other groups of model points as possible matches for that image group. Indexing is key to the fast implementation of the latter two methods above, because they all involve hypothesized matches between groups of image and model points, and indexing determines which groups of model points are possible matches for groups of image points. Search methods can also benefit by using indexing to determine which matches are most likely to be correct and using this information to guide the search.

Indexing systems typically require point groups to be of some minimum cardinality to perform their function correctly. A previous indexing system for indexing three-dimensional model groups from two-dimensional image groups [Clemens and Jacobs, 1991] required groups of size at least four and each group was represented on a two-dimensional surface in a four-dimensional table. By using a probabilistic method that allows false negatives (matches that are correct, but are not indexed) between image groups and model groups, we have designed a system which can index on groups of size three, and represents groups in only a single bucket in a two-dimensional look-up table.

The ability to index on groups of size three is very important. If there are n image points and m image points, then there are $O(n^k)$ image groups and $O(m^k)$ model groups of size k, so reducing the required group cardinality necessary reduces the number of groups to consider immensely. Several algorithms (e.g. [Huttenlocher and Ullman, 1990, Lowe, 1987]) use initial matches of three image points to three model points because this is the minimum number necessary to determine a finite set of transformations that bring the points into alignment. Indexing systems that require groups of larger than three points cannot generate ideal candidate matches for these algorithms.

In this work, we use the probabilistic peaking effect [Ben-Arie, 1990, Binford *et al.*, 1989, Burns *et al.*, 1990] to discriminate between likely matches and unlikely matches. The principle of the probabilistic peaking effect is that angles and ratios of distances between points in the model groups do not vary much as the viewpoint changes over much of the viewing sphere. This means that the probability density functions of these angles and ratios of distances of projected (image) points have a strong peak at the pre-projection (model) value. Binford *et al.* call such features 'quasi-invariants' because of their relative lack of variation with the change of viewpoint.

Let us call a set of image points hypothetically grouped for use in indexing the table an *image group*, and the model points hypothetically matched to them a *model group*. If each of the points in the image group is a result of the projection of its corresponding model group point then we will say that the two groups are in *actual correspondence*. The premise of our system is that the probabilistic peaking effect is a strong enough indicator of model feature values to eliminate the vast majority of model groups which are not in actual correspondence with a specific image group while keeping a significant percentage of those that are in actual correspondence.

Ben-Arie [1990] gives an equation to approximate the joint probability density function of features for model groups of size three, but this approximation does not vary explicitly with the angle of the model point group, only with the ratio of image angle to model angle, while the true effect varies over both quantities. We have recreated the experiments for determining the probabilistic peaking effect using the model angle as an additional parameter to achieve more accuracy modeling the effects of varying the model angle.

To incorporate the probabilistic peaking effect into a probabilistic indexing system, we use the joint probability density functions to determine what ranges of parameters a model group may have and still be likely to have generated a specific image group. This determines which groups are indexed. Our system can be extended to model and image groups larger than three points. This allows us to achieve better accuracy in discriminating between correct and incorrect matches, although with lower speedup.

In the following sections of this paper we shall give an overview of work on previous indexing systems and describe the probabilistic peaking effect in more detail. Section 4 will discuss how the probabilistic peaking effect can be used to build an indexing system. This is followed by a description of how the system can be extended to use image and model groups larger than three points. Section 6 gives indexing results on real images. Finally, we discuss how probabilistic indexing can be used to facilitate object recognition algorithms and other issues and conclude the paper.

2 Indexing for Object Recognition

Indexing systems for machine vision attempt to generate a single number (or at least a finite set of numbers) from an image or images that can be used to index a table and determine the set of models that could have generated the data. Ideally, one is able to represent the set of features comprising all or part of the object as a single number which remains the same regardless of transformation or projection. Such a number is called an *invariant*.

Invariants have been found for several types of model representations. For example, Lamdan *et al.* [1988] describe invariants for two-dimensional point sets (of size four or more) undergoing general three-dimensional affine transformations and orthographic projection. They represent a group of four points in a look-up table by the coordinates of the fourth point p_4 in a coordinate system with origin p_1 and unit axes $\overline{p_1p_2}$ and $\overline{p_1p_3}$. Model groups are then indexed by selecting three image points as a basis and using the relative coordinates of the remaining image points as keys into the index table. Voting is done to determine which objects might be present in the image. Forsyth *et al.* [1991] describe invariants for two-dimensional algebraic curves (e.g. conics) using the coefficients of the curve's equation. Differential invariants for general two-dimensional curves are given by Weiss [1992].

It has been proven that no invariants exist for single views of three-dimensional points sets [Burns *et al.*, 1990, Clemens and Jacobs, 1991] (for invariants from multiple views see Barrett *et al.* [1991].) Despite this, Clemens and Jacobs [1991] have shown that an indexing system for this problem can be built that (in the noiseless case) indexes exactly those groups that could have projected to a specific image group. This system requires groups to be of (at least) four points. It uses a four-dimensional index table and each group of four points must be represented over a two-dimensional space in this table. The requirement of four points per group means that there are $O(n^4)$ image groups and $O(m^4)$ model groups to consider. While this system achieves greater relative speedup by increasing the size of the point groups examined and the dimensionality of the index table, Clemens and Jacobs show that grouping is necessary for larger point groups to be of significant use, due to the larger number of groups found when the size of the point group is increased.

The four dimensions of the index table used by Clemens and Jacobs are as follows: the relative coordinates (x_4', y_4') of the orthographic projection of the fourth point using the projections of the first three as a basis (note that after projection these two parameters are the same representation used by Lamdan *et al.*,) the angle formed by the projections of the first three points, and the ratio of vector lengths of the projections of the first to second and first to third points (these are the same parameters used by Ben-Arie [1990] to demonstrate the probabilistic peaking effect.) These four parameters are invariant for any group of four points over translation, scaling, and rotation about the viewing direction. These parameters are not invariant over the remaining viewing parameters (i.e. viewing direction in the weak-perspective model,) thus they must represent each group from each viewing direction in their index table. Since the viewing direction has two degrees of freedom, this means that each group must be represented on a two-dimensional surface in the table.

Jacobs [1992] has shown that groups of four three-dimensional points can be indexed from two-dimensional data by representing each group as one-dimensional surfaces in two orthogonal two-dimensional tables. To determine which model groups may have projected to an image group, model groups are indexed in both of the tables. The intersection of the two sets of indexed groups then corresponds to the possible model groups.

Lamdan and Wolfson [1988] have extended their system such that it can also deal with the case of three-dimensional models and two-dimensional data. Their system accomplishes this by indexing on groups of five image points. While each model group is represented only once, each image group must index a line in the look-up table. Clemens and Jacobs show that this method does not take advantage of all of the constraints available and thus unnecessarily produces many false positive group matches. The requirement of groups of five points also means that there are $O(n^5)$ image groups and $O(m^5)$ model groups to consider.

We propose to use the probabilistic peaking effect to determine which model groups are likely to match specific image groups. The advantage of this method is that smaller image and model groups can be used (we require a minimum of three points per group) and each model group must be represented only once in the look-up table. This reduces the number of groups we must consider to $O(n^3)$ image groups and $O(m^3)$ model groups. The primary disadvantage is that we will not index all of the model groups in actual correspondence. While we will index many incorrect model groups, other indexing methods share this problem. (We will compare our method to others in Section 8.) In our method, each image group will index a small area of the look-up table which we will call a *cloud*. In the presence of error, every indexing method must either examine some non-infinitesimal volume of the index table for a particular image group or represent each model group in some non-infinitesimal volume of the index table, but most other indexing methods can be analyzed in the noiseless case as indexing an infinitesimal volume of the index table, while ours cannot.

3 The Probabilistic Peaking Effect

While it has been proven that there is no affine or projective invariant for three-dimensional point sets, it has been observed that there is a strong peaking effect for many angles and ratios of lengths in images at the values taken by the features in the model [Ben-Arie, 1990, Binford *et al.*, 1989, Burns *et al.*, 1990]. This information can be used to discard matches between image points and model points that have a small likelihood of actual correspondence.

We use the features defined by Ben-Arie [1990] to determine which groups are

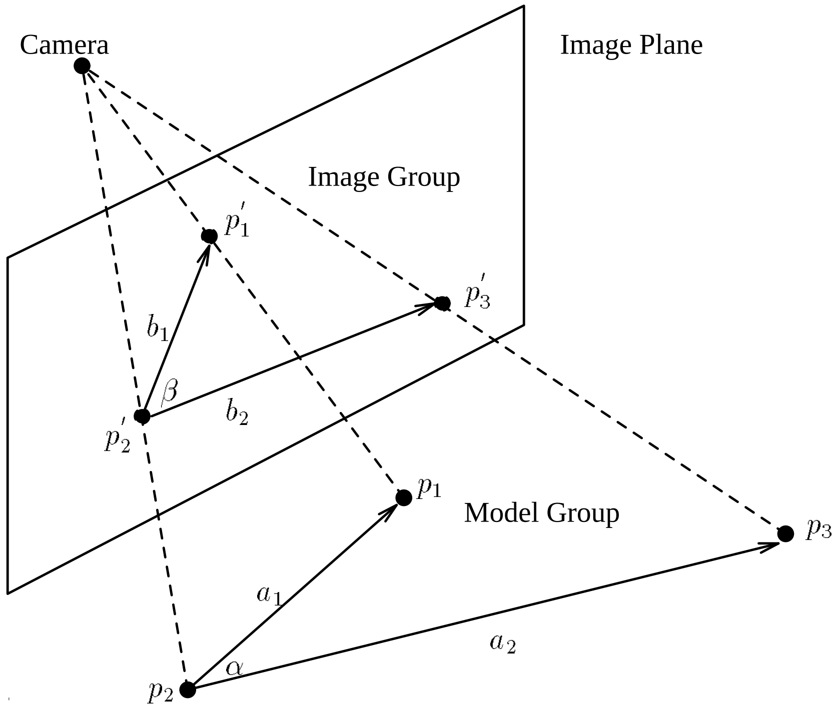


Figure 1: Projected Model Group

likely to match. These features are easy to determine for point sets of three ordered points. Let p_1, p_2, and p_3 be the points in the model group and p_1', p_2', and p_3', be the corresponding image points. Also, let α be the angle $\angle p_1p_2p_3$ and β be the angle $\angle p_1'p_2'p_3'$. Define the segment lengths as follows: $a_1 = |\overline{p_1p_2}|, a_2 = |\overline{p_2p_3}|, b_1 = |\overline{p_1'p_2'}|, b_2 = |\overline{p_2'p_3'}|$. See Figure 1. The features used are:

1. The angles formed by the points in the model (α) and in the image (β).

2. The ratios of the lengths of the segments ($\frac{a_1}{a_2}$ and $\frac{b_1}{b_2}$).

Ben-Arie gives an equation to approximate the probabilistic peaking effect as it varies over $\frac{\beta}{\alpha}$ and $\frac{b_1a_2}{b_2a_1}$. It should be noted that the peaking effect varies not only with the ratio $\frac{\beta}{\alpha}$ and $\frac{b_1a_2}{b_2a_1}$, but also with α (or alternately with α, β, and $\frac{b_1a_2}{b_2a_1}$). Ben-Arie's approximation of the joint probability density does not model this effect. To better model the probabilistic peaking effect, we have created probability histograms through numerical integration with the additional variable α. Like the experiments performed by Ben-Arie, we tessellated the viewing sphere and added the area of each tessellation in the table corresponding to the model angle α to the bucket corresponding to the image angle β and the logarithm of the ratio of ratio of lengths ($\log \frac{b_1a_2}{b_2a_1}$) from the viewing direction at the center of the tessellation. Ben-Arie uses buckets that vary uniformly with $\log \frac{\beta}{\alpha}$ to measure the angle variation. We use buckets varying uniformly in β because it has explicit bounds (0,180) and because we model the variation in

α explicitly. Since it is unclear how the objects in the images will be distributed with respect to distance from the camera, the orthographic projection was used in these experiments. Note that using the orthographic project, the probability density varies with $\frac{b_1 a_2}{b_2 a_1}$. Changing the pre-projection ratio of lengths $\frac{a_1}{a_2}$ has no effect on the probability density.

The result of these numerical integrations is an array of two-dimensional joint probability histograms, where Ben-Arie had a single joint probability density. Figure 2 shows the probability histograms in the noiseless case for $\alpha = 30$, $\alpha = 90$, $\alpha = 140$, and $\alpha = 160$. The x-axis is $\log \frac{b_1 a_2}{b_2 a_1}$, the y-axis is the image angle β, and the z-axis is the probability. As expected, the closer the model angle α is to 0 or 180 degrees, the stronger the peak.

To account for noise we have also generated the probability histograms with bounded noise ($\epsilon = 1.0 and 3.0$) added to the image parameters. The bounded noise model specifies that the true location of each image feature is within some distance ϵ of the measured location. The bounded noise model is used here to facilitate comparison against other systems, but any noise model can be dealt with in this fashion.

This method of accounting for noise should be adequate if we are dealing with images with approximately the same noise distribution. If we examine a number of images with different noise distributions we may not want to store several of the joint probability histograms, since they are large. An alternative would be to determine the possible ranges of the true values of the image features at run time from the observed values and the noise distribution. We would then determine which look-up table buckets must be examined using these ranges of values. This alternative has the disadvantage of slower run-time operation. Figure 3 shows the joint probability histograms for the case with noise ($\epsilon = 1.0$).

4 Probabilistic Indexing

The probabilistic peaking effect can be used to create a probabilistic indexing system to determine which model groups are most likely to have projected to specific image groups. The first step is to create a look-up table containing the model group information. The angle (α) and ratio of lengths ($\frac{a_1}{a_2}$) is determined for each model group in each model and the necessary information about these model groups is stored in the appropriate bucket in the table. This table is quantized in the same manner as the peaking effect probability histograms to facilitate indexing. Note that this table is two-dimensional and each model group is stored in a single bucket.

To determine which model groups are likely to have projected to an image group, we search the probability histograms described in the previous section. The parameters over which this search must vary are the angle α (this determines which histogram we examine) and the ratio $\frac{b_1 a_2}{b_2 a_1}$ within each histogram. We do not need to vary the angle β within each histogram because this is fixed by the image group angle. Since the probability of a particular set of image features is highest when the model values

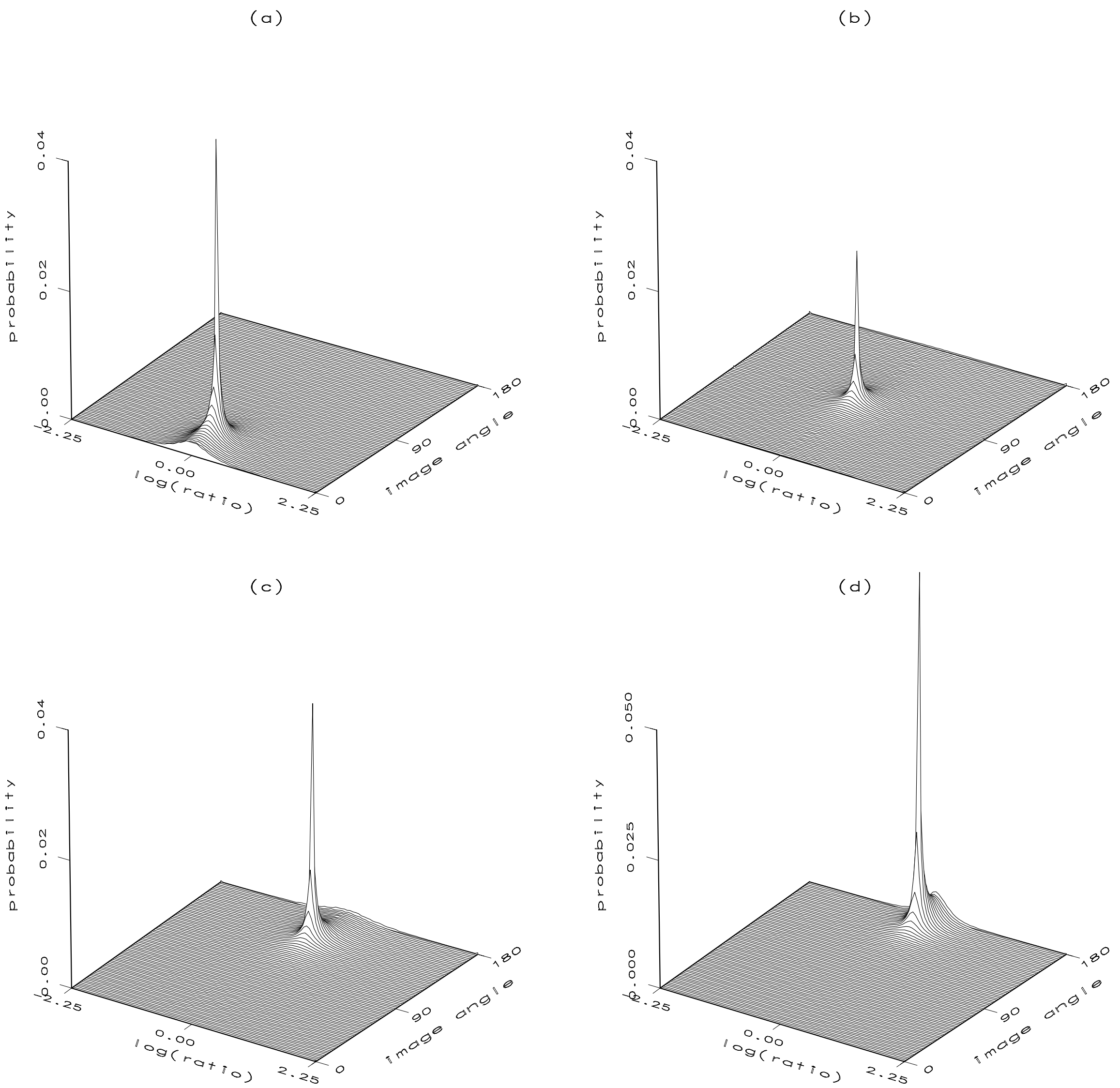


Figure 2: Joint probability histograms without noise. (a) $\alpha = 30°$ (b) $\alpha = 90°$ (c) $\alpha = 140°$ (d) $\alpha = 160°$

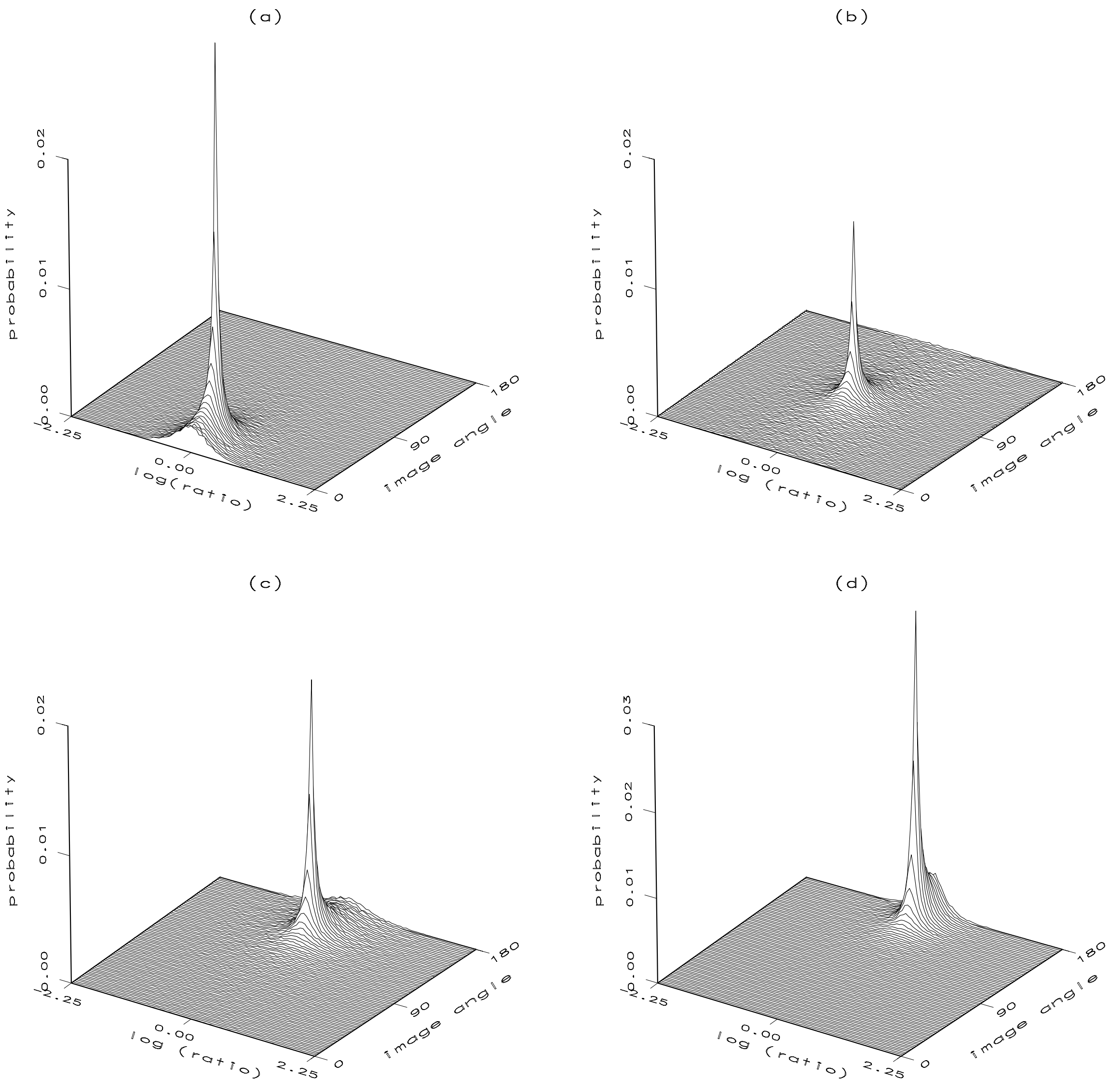


Figure 3: Joint probability histograms with noise. (a) $\alpha = 30°$ (b) $\alpha = 90°$ (c) $\alpha = 140°$ (d) $\alpha = 160°$

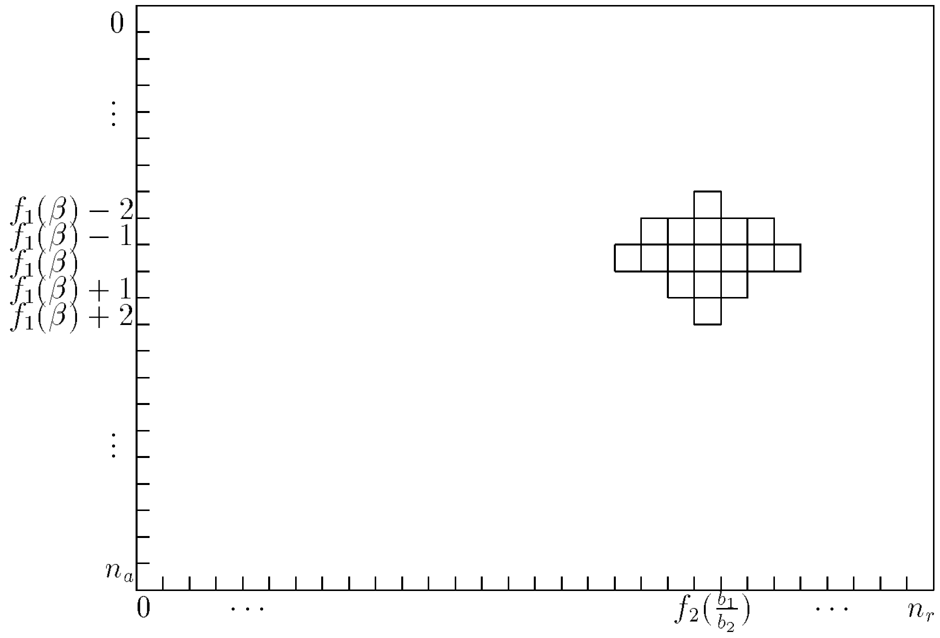


Figure 4: An example cloud of look-up table buckets.

are the same as the image features, we search outward from the bucket corresponding to the image feature values to determine which buckets in the look-up table we must examine. This search determines an area of buckets in the index table that we call a cloud. Each bucket in the cloud is be examined for model groups that may match this image group.

Let $f_1(\beta)$ be the row of the index table corresponding to the image angle and $f_2(\frac{b_1}{b_2})$ be the column corresponding to the image ratio of lengths. Figure 4 shows an example cloud in the look-up table. Note that it is centered at the bucket corresponding to the image feature values.

The extent of a cloud is determined as follows: for each angle α, we examine the row corresponding to the image group angle β and determine what range (if any) of ratios $\frac{b_1 a_2}{b_2 a_1}$ has a probability above a predetermined constant. (This constant is determined a priori to eliminate most groups not in actual correspondence, while keeping a large number of those that are. See below.) This provides the information to determine which buckets in the look-up table contain the model groups most likely to match this image group: for each α, we determine the range of ratios $\frac{a_1}{a_2}$ that should be examined in the look-up table from the range of $\frac{b_1 a_2}{b_2 a_1}$ determined as described above and $\frac{b_1}{b_2}$ from the image group. Each model group contained in these buckets is considered as a possible match for the current image group. We do not need to worry about noise in the index features when indexing because we have already accounted for it in the probabilistic peaking effect probability histograms.

Figure 5 illustrates the flow of information in the probabilistic indexing process.

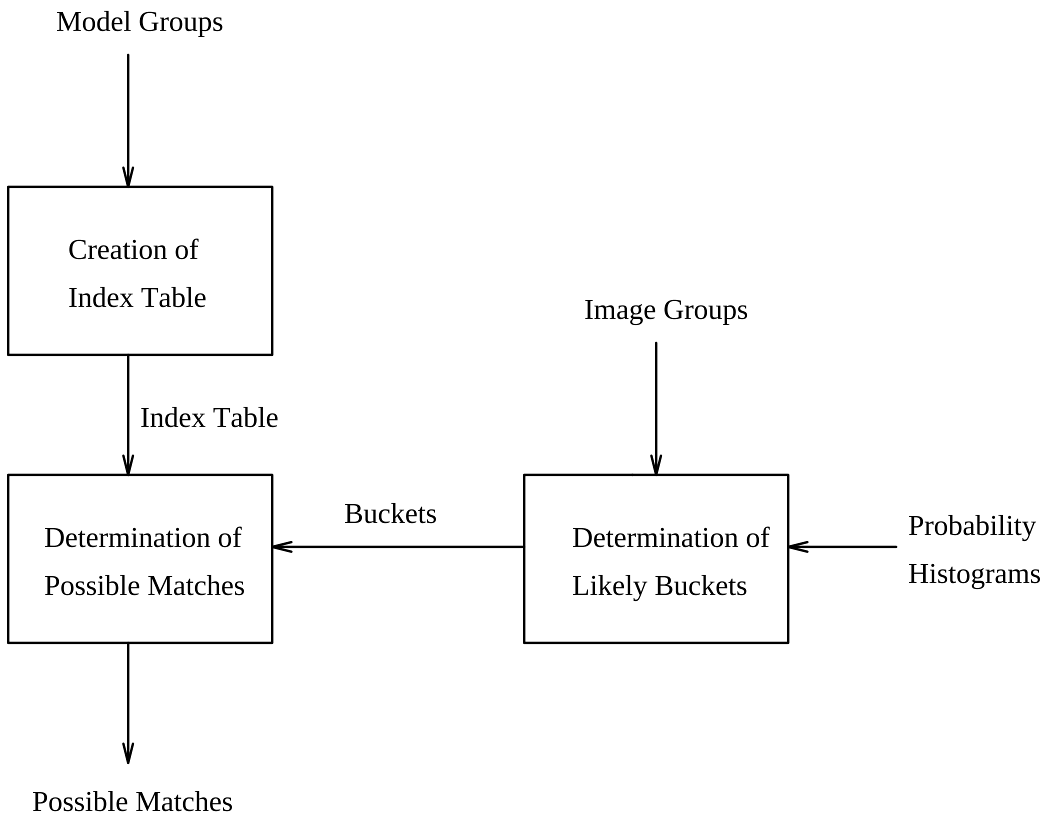


Figure 5: The indexing process.

The index table is created by storing the model groups in the appropriate buckets. Image groups and probability histograms are used to determine which buckets in the index table are most likely to contain the matching model group. These buckets are then examined in the index table to find possible matches for the image group.

Table 1 shows the percentage of total matches and matches in actual correspondence indexed for various probability thresholds as determined by experiments on objects of random three-dimensional points. In this table, K_t is the probability threshold used to determine if a bucket need be examined, p is the percentage of total matches found in the examined buckets and ρ is the percentage of matches in actual correspondence found in the examined buckets. The ratios of these values $\frac{\rho}{p}$ is the relative frequency of indexing correct and incorrect matches and $\frac{1}{p}$ is the speedup attained if we simply use these techniques to determine which matches are likely to be in actual correspondence in conjunction with an algorithm that requires hypothesized matches, such as alignment [Huttenlocher and Ullman, 1990]. The percentages were determined using random three-dimensional points, transformed by a random three-dimensional rotation and projected using the perspective projection. Bounded

K_t	p	ρ	$\frac{\rho}{p}$	$\frac{1}{p}$
.001	.0663	.468	7.06	15.08
.002	.0282	.335	11.88	35.46
.003	.0164	.265	16.16	60.98
.004	.0114	.226	19.82	87.72
.005	.0085	.193	22.71	117.65
.006	.0066	.173	26.21	151.52
.007	.0054	.159	29.44	185.19
.008	.0044	.140	31.82	227.27
.009	.0036	.128	35.56	277.78
.010	.0031	.115	37.10	322.58

Table 1: Elimination percentages for various peaking cutoffs with noise

noise ($\epsilon = 1.0$) was added to each of the feature coordinates.

We expect the probability of indexing a correct match will be better for observed image points compared to the random points used in our experiments. This is because model groups that appear in unlikely positions (i.e. such that they are highly foreshortened) are more likely to have one or more points occluded by the object itself, while our experiments assume no self-occlusion. We therefore expect groups of observed image points from real objects to produce a higher rate of indexing the correct group than random points.

If we know the prior probability distribution of image group features we can use the probabilities in Bayes' rule:

$$P(H|E) = \frac{P(H)P(E|H)}{P(E)}$$

where H is the hypothesis and E the evidence. Our hypothesis is that a particular image group is the result of the projection of a particular model image group. The evidence is the angles and ratios of the image group. So, $P(E|H)$ is given by the peaking effect probability histograms and $P(E)$ is given by the prior probabilities of image group parameters. We assume that each model group is equally likely to appear in the image, so the prior probability of our hypothesis $P(H)$ is the same for each case. Of course, if we had knowledge that models were not equally likely to appear in the image we could use it here.

The joint prior probability histogram of the image group parameters β and $\log \frac{b_1}{b_2}$ for feature points that are the result of model feature points in the database (and not random image points) can be determined by averaging the probabilistic peaking histograms for the set of model groups. For each random model group we add to the average the joint probability histogram for the correct α shifted on the ratio axis

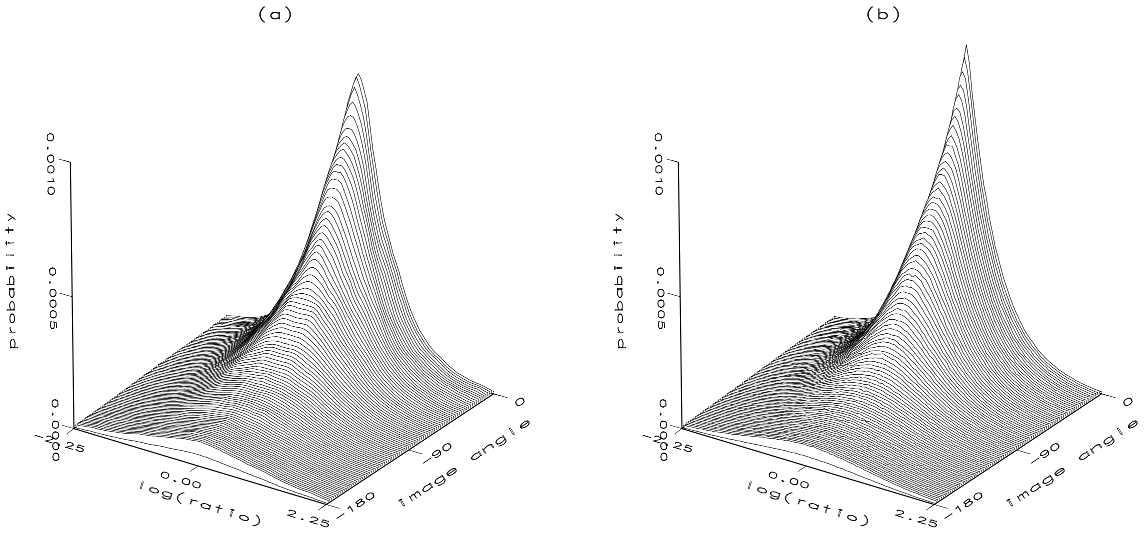


Figure 6: Prior probability histograms (a) projected random model points (b) random image points

by $\log \frac{a_2}{a_1}$. (A shift is required since the peaking histograms are for $\log \frac{b_1 a_2}{b_2 a_1}$ and we want the probability of $\log \frac{b_1}{b_2}$.) Again, these can be weighted if we know the prior probability of each model group appearing in the image.

This does not account for random extraneous points in the image. We can estimate the distribution of these points by examining the distribution of feature parameters for a large set of randomly selected image points. The prior probability histogram for image parameters for both model points and random points is shown in Figure 6. Since they are very close, we use the histogram for the projected model points as the prior probability histogram of the image group parameters for all image points.

We find that even among groups with high prior probability of matching (those that surpass the threshold, and thus are indexed,) matches in actual correspondence have, on average, considerably higher posterior probability. We'll call the expected posterior probability of a correct match that is indexed γ and the expected posterior probability of an incorrect indexed match g. Table 2 shows the values these take (neglecting the constant $P(H)$ term both have) for several indexing thresholds along with the ratio $\frac{\gamma}{g}$. Since the matches in actual correspondence have a considerably higher expected posterior probability, we can use the posterior probability to order the matches based on likelihood, if desired.

K_t	γ	g	$\frac{\gamma}{g}$
.001	31.35	6.08	5.15
.002	40.89	10.41	3.93
.003	48.88	13.86	3.53
.004	54.62	16.80	3.25
.005	60.40	19.32	3.13
.006	65.16	21.84	2.98
.007	68.36	24.08	2.83
.008	72.64	25.81	2.81
.009	75.67	27.72	2.73
.010	81.37	29.86	2.72

Table 2: Values of γ, g and $\frac{\gamma}{g}$ for various thresholds.

5 Using Larger Groups

Some algorithms require hypothesized point group matches of more than three points. Probabilistic indexing can be extended to accommodate these algorithms. This has the additional benefit of increasing our ability to discriminate between correct and incorrect matches. To incorporate larger groups into probabilistic indexing we must be able to index model groups of size k using keys determined from image groups of size k. This is accomplished by examining each subgroup of three points in the k point image groups in the manner of the previous section. If some predetermined constant number x_0 of the subgroups from a model group have high enough probability of matching, then the model group is considered a possible match. To determine if this is the case, we must index the look-up table with each combination of three points in the k point image group and determine which k point model groups are indexed at least x_0 times. The model groups are examined to ensure that each image point corresponds a model point consistently in the matches it was found in. That is, we don't want a particular model point to correspond to one image point when indexed by one subgroup and then correspond to a different image point when indexed by another subgroup and yet still be considered as a possible match.

Call p_0 and ρ_0 the values of p and ρ we had for indexing three point groups. Table 3 shows the new values of p and ρ for larger groups determined experimentally using random model groups and transformations for various values of p_0, ρ_0, k, and x_0. Since, for larger point groups ρ increases or stays about the same and p substantially decreases, increasing the size of groups used increases the ability of probabilistic indexing to discriminate between correct matches and incorrect matches. The price we pay for this accuracy is speed. Grouping processes will typically find more potential groups when the size of the group is increased, and for each group we must now

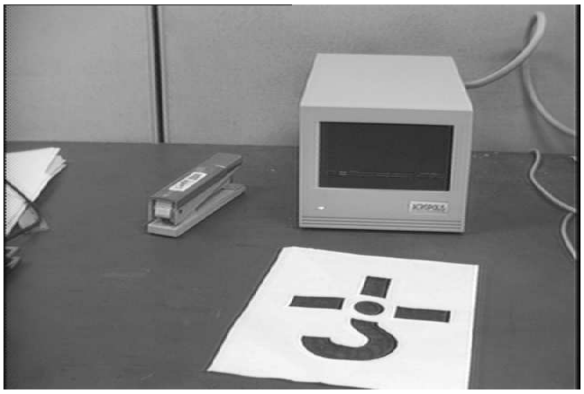
	ρ_0	p_0	ρ	p	$\frac{\rho}{p}$
$k=4$	.292	.0263	.289	.00788	36.68
$x_0=2$	.196	.0103	.176	.00172	102.33
	.155	.0059	.127	.00075	169.33
$k=5$	.313	.0286	.530	.00792	66.92
$x_0=3$	.204	.0113	.315	.00124	254.03
	.152	.0065	.206	.00039	528.21
$k=6$	.311	.0276	.736	.00823	89.43
$x_0=4$	.209	.0109	.515	.00114	451.75
	.153	.0062	.348	.00040	870.00

Table 3: Values of p and ρ for various parameters.

index $\binom{k}{3} = \frac{k!}{3!(k-3)!}$ subgroups of three points. This can be alleviated somewhat by bookkeeping techniques since there are at most $\binom{n}{3}$ total subgroups in the image, where n is the number of feature points in the image.

Clemens and Jacobs [1991] are able to increase the speedup of their system by increasing the size of the groups and the dimensionality of the index table because they are able to canonically order the points in each group. This means they don't need to test each of the $k!$ orderings of each group. They can canonically order their points since each representation of a model group in the index table is from a single viewpoint. This makes the implicit assumption that localization error will not disturb the image feature points enough to change the canonical ordering. In the general case, each of the $k!$ orderings must be stored in the index table. This method cannot be used with probabilistic indexing because we can't order the image points in a canonical manner (viewing the points from a different direction would generally lead to a different ordering.) This means we would have to examine each of the $k!$ orderings. In addition, we would need to store each of the $\begin{pmatrix} \binom{k}{3} \\ x_0 \end{pmatrix}$ combinations of subgroups that would indicate that a model group should be indexed separately in the index table. This extra cost would negate any extra speedup that increasing the dimensionality of our index table could produce.

We argue that using larger groups may not be as beneficial as Clemens and Jacobs claim. The larger the group a grouping process must find, the less likely all of the points in a group will arise from the same object (a point Clemens and Jacobs do not consider.) Any group of points that do not all arise from the same object is useless for indexing. This means that even though the speedup may be increased considerably by examining larger groups, a smaller percentage of the groups that are examined will be useful.



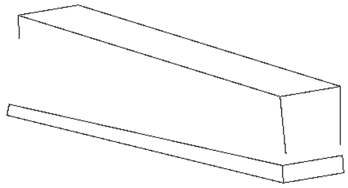


Figure 7: (a) An image used for testing (b) Stapler model with visible edges drawn in.

6 Results on Real Images

Probabilistic indexing has been tested on several real images. For these tests, model points on each of the objects were measured by hand. Several images of these objects were captured. Corners were determined by hand with the help of an edge detector. Figure 7.a shows an image used in the tests containing a disk, a stapler, and a hand rendering of a symbol from mythology. Figure 7.b displays an sample object model. The location of many of the feature points on the stapler are shown with the visible edges drawn in.

Table 4 gives the results of using the feature points from real images to index a database of 6 real objects. The average percentage of correct groups (ρ) and percentage of incorrect groups (p) that were indexed is shown from experiments using 5 images of the stapler, 3 images of the disk, and 3 images of the cross. Also given are the results showing how often random points indexed these model groups.

In each of the cases, the real feature point groups indexed the correct model group with frequency higher than was obtained for models of random points (see Table 1.) The frequency of indexing incorrect model groups was also slightly higher in many cases except for the stapler where it is substantially greater. The random image points indexed the real model groups with comparable, although higher, frequency than random image points indexed random model groups.

The rendering of the mythological symbol is a two-dimensional model. Figure 7.a shows the image of this object that we tested that had the most foreshortening. While the percentage of correct matches indexed was below the average for this image, we still indexed over 10% of the correct matches even when the probability threshold was high (K_t =.010).

Object	K_t	ρ	p
Stapler		.337	.0473
Disk	.002	.416	.0277
Cross		.597	.0261
Random		-	.0291
Stapler		.257	.0270
Disk	.004	.323	.0128
Cross		.459	.0124
Random		-	.0144
Stapler		.205	.0181
Disk	.006	.271	.0074
Cross		.382	.0078
Random		-	.0090
Stapler		.182	.0138
Disk	.008	.230	.0053
Cross		.326	.0057
Random		-	.0066
Stapler		.152	.0103
Disk	.010	.204	.0037
Cross		.282	.0043
Random		-	.0053

Table 4: Results of indexing table of real models.

7 Discussion

Probabilistic indexing should not be viewed as a method of adding randomization to the indexing problem, since the orientations from which each group is indexed are correlated. We rely on the fact that there are so many (approximately $\frac{m^3}{6}$) model groups that all viewing directions will have some model groups that are viewed in a likely orientation. If there is not a wide variety of orientation of the groups themselves, this may not be the case. In the extreme case, flat objects will have only a single orientation that all of the model groups share. Model groups from such objects will not be indexed correctly for many viewing directions, but we can easily determine which objects are flat or nearly flat prior to recognition time. For such objects we can either reduce the probability threshold or use special case techniques for flat [Lamdan *et al.*, 1988] or nearly flat [Arbter *et al.*, 1990] objects.

Contrasting probabilistic indexing to grouping techniques [Ahuja and Tuceryan, 1989, Huttenlocher and Wayner, 1992, Lowe, 1985, Mohan and Nevatia, 1992] may be useful. Grouping techniques determine sets of image features that are likely to come from the same object. These techniques can also be applied to models to determine which sets of points are likely to be found by grouping in the image. These techniques drastically lower the number of groups in the image and model that must be examined. Rather than examing groups of model points and groups of image points separately as grouping techniques do, probabilistic indexing examines matches between sets of model points and image points to determine which matches are most likely to be correct. Grouping can be used to predetermine which sets of model and image points are examined by the probabilistic indexing system to further reduce the number of matches that must be examined.

Probabilistic indexing techniques can be used to improve the performance of many current recognition systems. For example, Olson [1992] describes a method of speeding up algorithms that hypothesize matches of model groups to image groups (e.g. the alignment method [Huttenlocher and Ullman, 1990].) Error criteria are used to determine model groups that produce small error in the calculation of the transformation aligning the model and image points. The remaining model groups are not considered, since these methods depend on the accurate determination of this transformation. The probabilistic peaking effect is then used to determine which of these model groups are the most likely to match groups of image points found in an image. These techniques can be implemented efficiently using probabilistic indexing. The model groups that are found to cause little uncertainty in determination of the transformation are placed in the index table. Image groups are then used to index into the table to find likely matches. These techniques have been shown to speed up the alignment method by over two orders of magnitude, with little chance of missing a correct object for models that are not flat.

We note that, at present, probabilistic indexing is the only indexing system able to index on groups of three points, thus any other indexing system used with the alignment method (or any other algorithm using matches of three points) would require the algorithm to examine groups larger than the ideal size. This causes an increase in the computational complexity of the algorithm since there are $O(n^k)$ image groups and $O(m^k)$ model groups.

Algorithms that hypothesize matches and then use clustering in transformation space (e.g. Thompson and Mundy [1987] or Grimson *et al.* [1992]) can also benefit. The elimination of unlikely matches would not only speed up such algorithms, it would also help alleviate the significant problem of false positives [Grimson and Huttenlocher, 1990] because the ratio of groups in actual correspondence to those that aren't would increase significantly, making false positive peaks relatively less strong.

8 Conclusion

We have described an indexing system for use in solving the problem of recognizing three-dimensional objects in single two-dimensional images. The probabilistic peaking effect has been shown to be effective for use in indexing model groups undergoing general rigid transformations in three-dimensions from image group parameters in images generated using the perspective projection. Its use has allowed us to reduce the cardinality of the sets of image and model points necessary in an indexing system, while retaining the indexing speedup. The disadvantage to this system is that not all correct matches between image groups and model groups are indexed, but since a far higher percentage of correct matches than incorrect matches probabilistic indexing is usually very useful. Probabilistic indexing can be used as a pre-processing step for any algorithm that hypothesizes matches between groups of image and model points. By selecting only those matches that are likely to produce good results, probabilistic indexing can speed up and improve the performance of such algorithms considerably.

Acknowledgements

The author thanks Jitendra Malik for his guidance in all aspects of this research. This research has been supported by a National Science Foundation Graduate Fellowship to the author, NSF Presidential Young Investigator Grant IRI-8957274 to Jitendra Malik, and NSF Grant IRI-9114446.

References

[Ahuja and Tuceryan, 1989] N. Ahuja and M. Tuceryan. Extraction of early perceptual structure in dot patterns: Integrating region, boundary, and component gestalt. *Computer Vision, Graphics, and Image Processing*, 48:304–356, 1989.

[Arbter *et al.*, 1990] K. Arbter, W. E. Snyder, H. Burkhardt, and G. Hirzinger. Application of affine-invariant Fourier descriptors to recognition of 3-d objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):640–647, July 1990.

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

[Ballard, 1981] D. H. Ballard. Generalizing the Hough transform to detect arbitrary shapes. *Pattern Recognition*, 13(2):111–122, 1981.

[Barrett *et al.*, 1991] E. B. Barrett, P. M. Payton, N. N. Haag, and M. H. Brill. General methods for determining projective invariants in imagery. *Computer Vision, Graphics, and Image Processing*, 53(1):46–65, January 1991.

[Ben-Arie and Meiri, 1987] J. Ben-Arie and A. Z. Meiri. 3d objects recognition by optimal matching search of multinary relations graphs. *Computer Vision, Graphics, and Image Processing*, 37(3):345–361, March 1987.

[Ben-Arie, 1990] J. Ben-Arie. The probabilistic peaking effect of viewed angles and distances with application to 3-d object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(8):760–774, August 1990.

[Binford *et al.*, 1989] T. O. Binford, T. S. Levitt, and W. B. Mann. Bayesian inference in model-based machine vision. In L. N. Kanal, T. S. Levitt, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 3*, pages 73–95. Elsevier Science Publishers B.V., North-Holland, 1989.

[Burns *et al.*, 1990] J. B. Burns, R. Weiss, and E. M. Riseman. View variation of point set and line segment features. In *Proceedings of the DARPA Image Understanding Workshop*, pages 650–659, 1990.

[Cass, 1990] T. A. Cass. Feature matching for object localization in the presence of uncertainty. In *Proceedings of the International Conference on Computer Vision*, pages 360–364, 1990.

[Clemens and Jacobs, 1991] D. T. Clemens and D. W. Jacobs. Space and time bounds on indexing 3-d models from 2-d images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(10):1007–1017, October 1991.

[Faugeras and Hebert, 1986] O. D. Faugeras and M. Hebert. The representation, recognition, and locating of 3-d objects. *International Journal of Robotics Research*, 5(3):27–52, 1986.

[Forsyth *et al.*, 1991] D. Forsyth, J. L. Mundy, A. Zisserman, C. Coelho, A. Heller, and C. Rothwell. Invariant descriptors for 3-d object recognition and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(10):971–991, October 1991.

[Grimson and Huttenlocher, 1990] W. E. L. Grimson and D. P. Huttenlocher. On the sensitivity of the Hough transform for object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(3):255–274, March 1990.

[Grimson *et al.*, 1992] W. E. L. Grimson, D. P. Huttenlocher, and T. D. Alter. Recognizing 3d objects from 2d images: An error analysis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 316–321, 1992.

[Grimson, 1990] W. E. L. Grimson. *Object Recognition by Computer: The Role of Geometric Constraints.* MIT Press, 1990.

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

[Huttenlocher and Wayner, 1992] D. P. Huttenlocher and P. C. Wayner. Finding convex edge groupings in an image. *International Journal of Computer Vision*, 8(1):7–27, 1992.

[Jacobs, 1992] D. W. Jacobs. Space efficient 3d model indexing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 439–444, 1992.

[Lamdan and Wolfson, 1988] Y. Lamdan and H. J. Wolfson. Geometric hashing: A general and efficient model-based recognition scheme. In *Proceedings of the IEEE Conference on Robotics and Automation*, pages 238–249, 1988.

[Lamdan *et al.*, 1988] Y. Lamdan, J. T. Schwartz, and H. J. Wolfson. Object recognition by affine invariant matching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 335–344, 1988.

[Lowe, 1985] D. G. Lowe. *Perceptual Organization and Visual Recognition.* Kluwer Academic, 1985.

[Lowe, 1987] D. G. Lowe. Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence*, 31:355–395, 1987.

[Mohan and Nevatia, 1992] R. Mohan and R. Nevatia. Perceptual organization for scene segmentation and description. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(6):616–635, June 1992.

[Olson, 1992] C. F. Olson. Fast alignment by eliminating unlikely matches. Technical Report UCB//CSD-92-704, University of California at Berkeley, November 1992.

[Thompson and Mundy, 1987] D. W. Thompson and J. L. Mundy. Three-dimensional model matching from an unconstrained viewpoint. In *Proceedings of the IEEE Conference on Robotics and Automation*, pages 208–220, 1987.

[Weiss, 1992] I. Weiss. Noise resistant projective and affine invariants. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 115–121, 1992.