

Probabilistic Indexing: A New Method of Indexing 3D Model Data from 2D Image Data

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

Abstract

Recent research has indicated 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 general 3-D objects from single 2-D images require groups of four points to generate a key into the index table and each model group requires many entries in the table.

I present a system that is capable of indexing using groups of three points by taking advantage of the probabilistic peaking effect [1]. Each model group need only be represented at one point in the index table. The ability to index using groups of three points means that there are many fewer model groups and image groups to consider, but to be able to index using groups of three points, false negatives matches must be allowed. We can withstand these false negatives by examining information from multiple groups.

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. 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 can be the key to the fast implementation of algorithms that rely on hypothesized matches between groups of image and model points, such as alignment and pose clustering.

Indexing systems typically require point groups to be of some minimum cardinality to perform their function correctly. Previous indexing systems for indexing general three-dimensional model groups undergoing rigid transformations from two-dimensional image groups [4] required groups of size at least four and each group was represented over a two-dimensional surface of 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, I 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 important. If there are n image points and m model 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 greatly reduces the number of groups to consider. Several algorithms 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, I use the probabilistic peaking effect [1, 2, 3] 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.

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 my system is that the probabilistic peaking effect is a strong enough indicator of model feature values to eliminate the vast majority 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.

In the following sections of this paper, I 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. Section 5 gives indexing results on real images. Finally, I discuss some interesting issues and conclude the paper.

2 Indexing for Object Recognition

In general, the goal of feature indexing systems is to determine which sets of model features may have projected under certain noise assumptions to specific sets of image features. These systems typically generate a vector of parameters that describe some key facets of a subset of an object's features. Ideally, these parameters do not change as the object is transformed and projected onto the image. In this case the parameters are called *invariants*.

Once invariants have been found, an index table can be created by discretizing the space of the invariant parameters. Model features are then binned in the table at the locations corresponding to their parameters. At recognition time, the parameters associated with the image features can be used to look up the model features in the index table that may correspond to them. In this paper, I will concentrate on the problem of indexing small sets of feature points.

Invariants have been found for several types of model representations. For example, Lamdan *et al.* [7] describe invariants for two-dimensional point sets (of size four or more) undergoing general three-dimensional affine transformations and orthographic projection. Algebraic and differential invariants for two-dimensional curves were first discussed by Weiss [12]. Forsyth *et al.* [5] used algebraic invariants of conics to recognize objects.

It has been proven that no invariants exist for single views of general three-dimensional points sets [3, 4]. While Rothwell *et al.* [11] demonstrate that invariants

exist for some constrained classes of three-dimensional point sets, I will not consider such constrained classes in this paper.

Despite this result on the lack of an invariant for three-dimensional point sets, Clemens and Jacobs [4] have shown that an indexing system for the general 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. The requirement of four points per group means that there are $O(n^4)$ image groups and $O(m^4)$ model groups to consider. This system is able to achieve greater speedup relative to the total number of groups by increasing the size of the point groups examined and size of the index table, but the overall time required increases due to the larger number of total groups, except when an accurate grouping method can select groups of image points that come from the same object.

I 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 (a minimum of three points per group are required) 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.

3 The Probabilistic Peaking Effect

While it has been proven that there is no affine or projective invariant for general three-dimensional point sets, it has been observed that there is a strong peaking effect in the probability densities of many angles and ratios of lengths in images at the values taken by the features in the model [1, 2, 3]. For example, assuming that every viewing direction is equally likely, the probability density of an angle formed by three points in an image has a strong peak at the actual angle formed by the points in real space. This means that there is a large range of viewing directions over which the angle formed by these points in these image changes by a small amount. Binford *et al.* [2] call such features 'quasi-invariants' because of their relative lack of variation with the change of viewpoint. This information can be used to discard matches between groups of image points and model points that

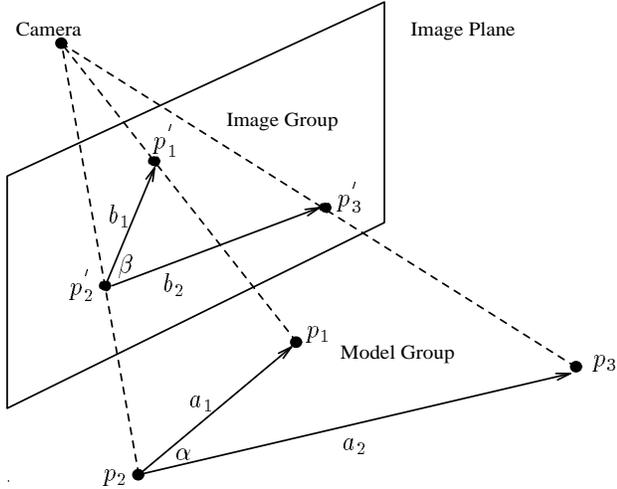


Figure 1: An example of a model group projected onto the image plane using the perspective projection.

have a small likelihood of being in actual correspondence.

The values used to determine which feature groups are likely to match in this system are determined as follows. 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 $\angle p_1 p_2 p_3$ and β be the angle $\angle p'_1 p'_2 p'_3$. Define the segment lengths as follows: $a_1 = |p_1 p_2|$, $a_2 = |p_2 p_3|$, $b_1 = |p'_1 p'_2|$, $b_2 = |p'_2 p'_3|$. 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 [1] gives an equation to approximate the probabilistic peaking effect as it varies over $\frac{\beta}{\alpha}$ and $\frac{b_1 a_2}{b_2 a_1}$. It should be noted that the peaking effect varies not only with the ratio $\frac{\beta}{\alpha}$ and $\frac{b_1 a_2}{b_2 a_1}$, but also with α (or alternatively with α, β , and $\frac{b_1 a_2}{b_2 a_1}$). Ben-Arie's approximation of the joint probability density does not model this effect. To better model the probabilistic peaking effect, I have created probability histograms, with the additional variable α , through numerical integration. These experiments tessellated the viewing sphere and added the area of each tessellation to the bucket corresponding to the image angle β and the logarithm of the ratio of lengths ($\log \frac{b_1 a_2}{b_2 a_1}$) from the viewing direction at the center of the tessellation. 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.

The result of these numerical integrations is an array of 2D joint probability histograms, where we previously had a single such 2D joint probability density. To account for noise, I have generated the probability histograms with bounded noise ($\epsilon = 1.0$ and 3.0) added to the image parameters. This method averages the effect of noise values on groups at different scales. Figure 2 shows the joint probability histograms for $\epsilon = 1.0$.

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 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 called a *cloud*. Each bucket in the cloud is examined for model groups that may match the current image group.

Let $r(\beta)$ be the row of the index table corresponding to the image angle and $c(\frac{b_1}{b_2})$ be the column corresponding to the image ratio of lengths. Figure 3 shows an example cloud in the look-up table. Note that it is centered at the bucket corresponding to the image feature values.

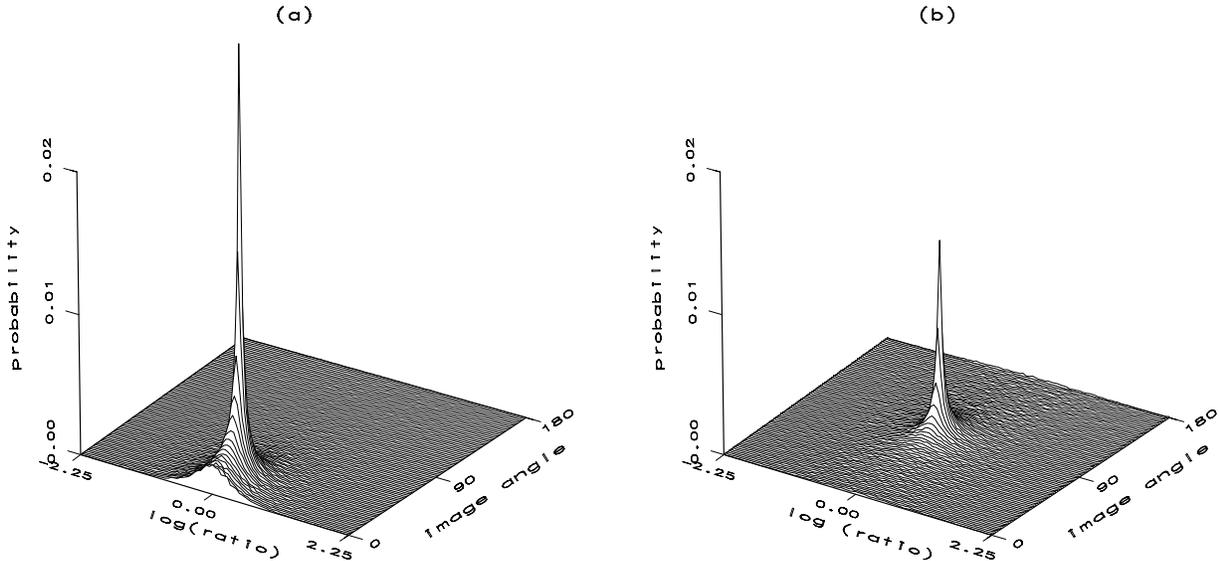


Figure 2: The joint probability histograms describing the probabilistic peaking effect for two values of model angle α with noise ($\epsilon = 1.0$). The x -axis is the image angle β . The y -axis is the logarithm of the ratio of lengths $\log \frac{b_1 a_2}{b_2 a_1}$. The z -axis is the probability. (a) $\alpha = 30^\circ$ (b) $\alpha = 90^\circ$.

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 image features when indexing because we have already accounted for it in the probabilistic peaking effect probability histograms.

Table 1 shows the fraction of total matches and matches in actual correspondence indexed for various probability thresholds as determined by experiments on objects of random three-dimensional points. These experiments transformed the models by a random three-dimensional rotation and projected them using the perspective projection. Bounded noise ($\epsilon = 1.0$) was added to each of the feature coordinates.

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: Fraction of correct and incorrect matches examined for various peaking parameter cutoffs with noise ($\epsilon = 1.0$). K_t is the the probability threshold used to determine if a match is eliminated, p is the fraction of incorrect matches examined, ρ is the fraction of of correct matches examined, $\frac{\rho}{p}$ is the relative frequency of indexing correct and incorrect matches and $\frac{1}{p}$ is the speedup attained if we simply used these techniques to determine matches that are likely to be in actual correspondence in conjunction with an algorithm that hypothesizes matches, such as alignment.

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) has been determined by averaging the probabilistic peaking histograms for the set of model groups. For each random model group, we averaged the joint probability histogram for the correct α shifted on the ratio axis 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 $\log \frac{b_1}{b_2}$.)

Using this prior probability distribution we can use Bayes' rule to determine the posterior probability of each match being in actual correspondence:

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

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. I have assumed 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 model groups were not equally likely to appear in the image we could use it here.

I have found that even among groups that are indexed, matches in actual correspondence have, on average, considerably higher posterior probability. Since the matches in actual correspondence have higher expected posterior probability, we can use the posterior probability to order the matches based on likelihood, if desired. In fact, we could easily use this probability as the indexing score, but this introduces a bias such that different correct groups are not equally likely to be indexed.

5 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 with the help of an edge detector.

Table 2 gives the results of using the feature points from real images to index a database of 6 real objects. The average fraction of correct groups (p) and fraction of incorrect groups (q) that were indexed is shown from experiments using 5 images of a stapler,

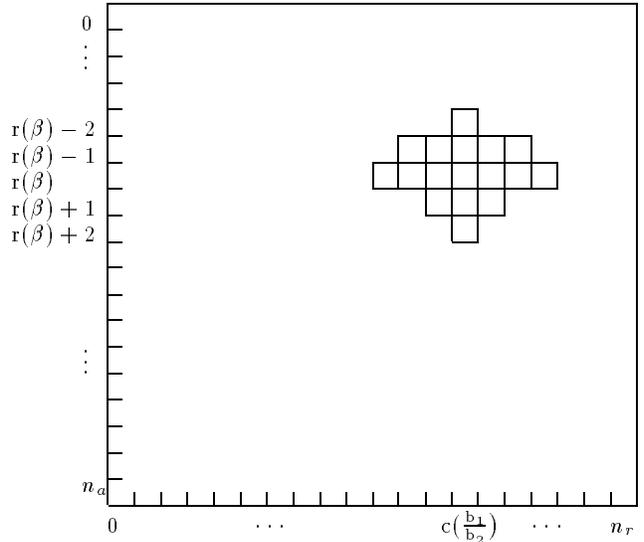


Figure 3: An example cloud of buckets in the index table. The hash marks along the x -axis represent the discretization of the possible values of $\frac{b_1}{b_2}$. The hash marks along the y -axis represent the discretization of the possible values of β . The marked bins are the bins that were found to have large enough probability of holding a correct match (i.e. the cloud.)

3 images of a disk, and 3 images of a planar symbol I'll call cross. Also shown are results on how often random points indexed groups from these models.

Comparing these results to those on synthetic data (see Table 1,) we see that in each of the cases, the real feature points indexed the correct model group with frequency higher than was obtained for correct groups from models of random points. The frequency of indexing incorrect model groups was also slightly greater in many cases, except for the stapler where it is substantially greater. The random incorrect image points indexed the real model groups with comparable frequency to indexing incorrect random model groups. These experiments indicate that the performance of the probabilistic indexing system on real images is similar to the results obtained for synthetic data.

6 Discussion

Probabilistic indexing should not be viewed as using randomization to benefit the indexing problem, since the orientations from which each group is correctly indexed are correlated. We rely on the fact

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

Table 2: Results of indexing using real objects and images: K_t is the probability threshold, ρ is the fraction of correct matches indexed, and p is the fraction of incorrect matches indexed.

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 and use special case techniques for such objects. In the experiments of the previous section the cross is a planar object. Even in the most foreshortened image tested, over 10% of the correct matches were indexed even when the threshold was high ($K_t = .010$), with only 0.43% of the incorrect matches indexed.

Probabilistic indexing techniques can be used to improve the performance of many current recognition systems. Algorithms that hypothesize matches of model groups to image groups are particularly good candidates for speedup. For example, Olson [8] describes a method of speeding up the alignment method [6] using probabilistic indexing. 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 the alignment method depends on the accurate determination of this transformation. Probabilistic indexing 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 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. Probabilistic indexing can also be used to select basis matches in the geometric hashing system [7] or select candidates to cluster in a pose clustering system (e.g. [10]).

Note that probabilistic indexing is the only indexing system able to index three-dimensional data using groups of three points. Thus, any other indexing system used to generate candidate basis matches for 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 if we examine points groups of size k , there are $O(n^k)$ image groups and $O(m^k)$ model groups.

Clemens and Jacobs [4] have argued that the use of large groups of feature points is beneficial when combined with grouping techniques that determine sets of feature points that are likely to come from the same object. Their reasoning is as follows. As the size of the point groups is increased, the relative speedup increases, but so does the total number of groups. When grouping is introduced, the growth of the number of groups is limited, but the indexing speedup is maintained. Thus using large groups will be good if accurate grouping is possible.

I argue that using larger groups will not be as beneficial as claimed. 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.

Finally, probabilistic indexing can be extended to point groups of larger than three points, if desired. This enables this system to do a better job discriminating between correct and incorrect matches, but would require more time. See [9] for details. It

should be noted, though, that as the size of the groups of points increases, probabilistic indexing fares worse than previous indexing algorithms. Thus, point groups of size three or four is optimal for probabilistic indexing.

7 Conclusion

I 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 an indexing speedup. The disadvantage to this system is that not all correct matches between image groups and model groups are indexed, but probabilistic indexing is useful since a far higher percentage of correct matches than incorrect matches are indexed. 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 advice on 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 Materials Handling Grant IRI-9114446.

References

- [1] 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.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] D. P. Huttenlocher and S. Ullman. Recognizing solid objects by alignment with an image. *International Journal of Computer Vision*, 5(2):195–212, 1990.
- [7] 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.
- [8] C. F. Olson. Fast alignment using probabilistic indexing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 387–392, 1993.
- [9] C. F. Olson. Probabilistic indexing: Recognizing 3d objects from 2d images using the probabilistic peaking effect. Technical Report UCB//CSD-93-733, Computer Science Division, University of California at Berkeley, May 1993.
- [10] C. F. Olson. Time and space efficient pose clustering. Technical Report UCB//CSD-93-755, Computer Science Division, University of California at Berkeley, July 1993.
- [11] C. A. Rothwell, D. A. Forsyth, A. Zisserman, and J. L. Mundy. Extracting projective structure from single perspective views of 3d point sets. In *Proceedings of the International Conference on Computer Vision*, pages 573–582, 1993.
- [12] I. Weiss. Projective invariants of shapes. In *Proceedings of the DARPA Image Understanding Workshop*, pages 1125–1134, 1988.