

Fast Alignment by Eliminating Unlikely Matches

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

October 30, 1992

Abstract

The alignment method [Huttenlocher and Ullman, 1990] is a model-based object recognition technique that determines possible object transformations from three hypothesized matches of model and image points. In the absence of grouping, the alignment method must examine each possible matching of three model points with three image points. Thus, if m is the number of model features and n is the number of image features, this method requires $O(m^3n^3)$ transformations to be computed. Each of these transformations must then be tested to determine whether it is correct, a time consuming step itself. For images and/or models with many features, the running time of the alignment method is not satisfactory, even in the presence of current grouping techniques.

This paper presents methods of reducing the number of matches that must be examined. The techniques we describe are: 1) using the probabilistic peaking effect [Ben-Arie, 1990] to eliminate unlikely matches, 2) examining the algorithm used to produce the transformation and eliminating model groups likely to produce large errors, 3) eliminating image groups likely to produce large errors. Results are presented that show we can achieve a speedup of over two orders of magnitude faster while still finding the best transformation.

1 Introduction

Model-based object recognition is a class of techniques to determine the location and pose of objects in images using a catalog of models of rigid objects. Typically, model-based systems hypothesize the existence of an object in the image and then perform some search to determine correspondences between image features and model features. The presence of this object is then verified while determining its location and pose. Some systems allow indexing a database with image feature statistics to reduce the number of objects that must be examined. Model-based object recognition techniques may recognize two- or three-dimensional objects in two- or three-dimensional image data.

The alignment method [Huttenlocher and Ullman, 1990] is a model-based object recognition technique for recognizing three-dimensional objects from a single view in two-dimensional images. For each model in the database, triples of image points are matched with triples of model points. For each match, the affine transformation that brings them into alignment is determined. The remaining model points (and/or other model features) are then transformed according to this transformation and compared against the remaining image points (and/or other image features) to verify the correctness of the transformation.

If every match of three model points to three image points is considered, the alignment method is $O(m^4 n^3 \log n)$ due to a $O(m \log n)$ verification step, where m is the number of model features and n is the number of image features. Huttenlocher and Ullman propose techniques that in some cases can reduce this to $O(m^3 n^2 \log n)$. It is clear that unless these time bounds have a very small leading constant, the running time will be considerable, since models with at least 20 features and images with over 50 features are common.

If a model object is present in the image, it is likely that a substantial number triples of model points can be detected. In the best case, only one of these triples needs to be found to recognize the object. If all image triples are examined, then much extra work is being done that is not necessary. Even if we can stop once a close enough match has been found, ordering the matches based on some likelihood of a good match can reduce the number of matches examined considerably. Grouping techniques determine sets of image points or lines that are likely to have resulted from the same object. See for example Lowe [1985], Huttenlocher and Wayner [1992], or Mohan and Nevatia [1992]. These techniques can be used to reduce the number image groups and model groups that are examined while retaining the important groups.

Instead of examining groups of model and image features separately, this paper addresses a method of greatly reducing the number of matches between groups of model and image features that must be examined by eliminating matches that are unlikely to produce correct transformations. These techniques can be used together with grouping and other methods to further reduce the number of matches that must be examined. The techniques that are used to eliminate matches are based on the following two principles:

1. The probability density functions of angles and distance ratios in images peak strongly at the pre-projection (model) value [Ben-Arie, 1990, Binford *et al.*, 1989, Burns *et al.*, 1990].

2. Matches that produce a transformation with a large uncertainty are unlikely to result in a good correspondence between the model and the image and are less likely to result in the verification routine determining that the object is present in the image. Examples of the use of similar principles can be found in Mundy *et al.* [1988] and Costa *et al.* [1989].

These techniques have been implemented for the alignment method and a speedup of over two orders of magnitude has been achieved while still finding a correct transformation.

Algorithms other than the alignment method will also benefit from these techniques. For example, algorithms that use clustering in transformation space (e.g. Thompson and Mundy [1987]) can eliminate matching groups that are unlikely to be correct. Aside from speeding up the algorithm, this elimination will also help alleviate the problem of false positives (which are a significant for clustering algorithms when faced with cluttered images [Grimson and Huttenlocher, 1990]) because the ratio of correct matches to incorrect matches will increase and thus random peaks will be relatively less strong.

In the next section, we will discuss the alignment method in more detail and summarize the algorithm. Section 3 will discuss the probabilistic peaking effect and how we use it to eliminate many matches between groups of image points and model points. Sections 4-6 describe how we determine which matches result in transformations with large uncertainties. Section 7 will present our experimental results including results on the efficacy of the affine approximation to the perspective projection. Section 8 discusses the techniques and results and includes an analysis of the speedup produced under various conditions. Finally, section 9 presents our conclusions.

2 The Alignment Method

The alignment method [Huttenlocher and Ullman, 1987, 1988, 1990] is a model-based technique for recognizing rigid three-dimensional objects from a single two-dimensional image. The premise of the alignment method is that a unique (up to a reflection) affine transformation between the model and image of the model can be found by matching three model points with three image points. This transformation is given by $p_i = \Pi(sR(p_m - p_{m_0})) + p_{i_0}$. Here, p_m is a model point and p_i is its corresponding image point, p_{m_0} and p_{i_0} are relative offsets, s is a scale factor, R is a three-dimensional rotation and $\Pi(\cdot)$ is the orthographic projection.

Let us call the set of model points being matched (three points per group for alignment) the *model group* and the image points hypothetically matched to them the *image 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 the two groups are in *actual correspondence*. For the rest of this paper we will consider a single object model. In practice, each model must be examined separately in the alignment method.

If the model is present in the image and we consider all possible matches between three model points and three image points it is very likely that many matches will be in actual correspondence.

The transformations determined from the matches in actual correspondence should then be close to the correct transformation. Each transformation computed must be verified against the image to determine if it is valid.

It is not advisable to examine each combination of three image points and three model points. If there are n image points and m model points, the entire algorithm would require $O(m^4 n^3 \log n)$ operations, since the verification routine requires $O(m \log n)$ operations. Huttenlocher and Ullman have used proposed various techniques to lower the complexity of the algorithm. In Huttenlocher and Ullman [1990], virtual points found by using the orientations at two model and image points are used to reduce the complexity to $O(m^3 n^2 \log n)$ operations.

The algorithm can be summarized as follows:

1. For each group of model points, rotate and translate the group such that the first point lies at the origin and the other points lie in the $x - y$ plane. This step is performed off-line for each model group and the new coordinates for the second and third points, b'_m and c'_m , are stored, as is a_m , which is the relative offset p_{m0} in the transformation described above.
2. For each image group (a_i, b_i, c_i) , solve for the 2x2 linear transformation matrix L , that brings the points into alignment in two dimensions, given by the following system of equations:

$$Lb'_m = b'_i \qquad Lc'_m = c'_i$$

where

$$b'_i = b_i - a_i \qquad c'_i = c_i - a_i$$

3. Determine the 3x3 linear transformation matrix sR^+ (due to the reflective ambiguity sR^- also exists and is described below) that maps the model onto the image, given by:

$$sR^+ = \begin{bmatrix} l_{11} & l_{12} & (c_2 l_{21} - c_1 l_{22})/s \\ l_{21} & l_{22} & (c_1 l_{12} - c_2 l_{11})/s \\ c_1 & c_2 & (l_{11} l_{22} - l_{21} l_{12})/s \end{bmatrix}$$

where

$$w = l_{12}^2 + l_{22}^2 - l_{11}^2 - l_{21}^2 \qquad q = l_{11} l_{12} + l_{21} l_{22}$$

$$c_1 = \sqrt{\frac{1}{2}(w + \sqrt{w^2 + 4q^2})} \qquad c_2 = -q/c_1$$

$$s = \sqrt{l_{11}^2 + l_{21}^2 + c_1^2}$$

4. Execute a verification procedure to determine if the transformation is correct. The transformed model points can be found by $p_i = \Pi(sR^+(p_m - a_m)) + a_i$ where p_m is the model point, p_i is its transformed location in the image, and sR^+ , a_m , and a_i are as given above. Since there is a reflective ambiguity, two transformations must be examined. The second, sR^- , is the same as sR^+ , except that the terms at positions (1, 3), (2, 3), (3, 1), and (3, 2) in the matrix are negated.

Note that the affine transformation is used in this algorithm as an approximation to the full perspective projection, and is valid only when the distance to the object is large compared to the size of the object in the z direction (after transformation) [Thompson and Mundy, 1987]. Experiments determining when the approximation is valid in practice are detailed in Section 7.

3 The Probabilistic Peaking Effect

While it has been proven that there is no invariant for three-dimensional point sets seen from a single view [Burns *et al.*, 1990, Clemens and Jacobs, 1991] (for an example of an invariant for three-dimensional point sets seen from multiple views see Barrett *et al.* [1991]), 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 features are unlikely 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 $\angle p_1 p_2 p_3$ and β be the angle $\angle p'_1 p'_2 p'_3$. Define the segment lengths as follows: $a_1 = |\overline{p_1 p_2}|$, $a_2 = |\overline{p_2 p_3}|$, $b_1 = |\overline{p'_1 p'_2}|$, $b_2 = |\overline{p'_2 p'_3}|$. Figure 1 gives an illustration. 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 finds the joint probability density to be well approximated by:

$$p_{12}(z, t) = (x_1 e^{-x_2|z|} + x_3 e^{-x_4|z|})(x_5 e^{-x_6|t|} + x_7 e^{-x_8|t|}) + (x_9 e^{-x_{10}u} + x_{11} e^{-x_{12}u})$$

where

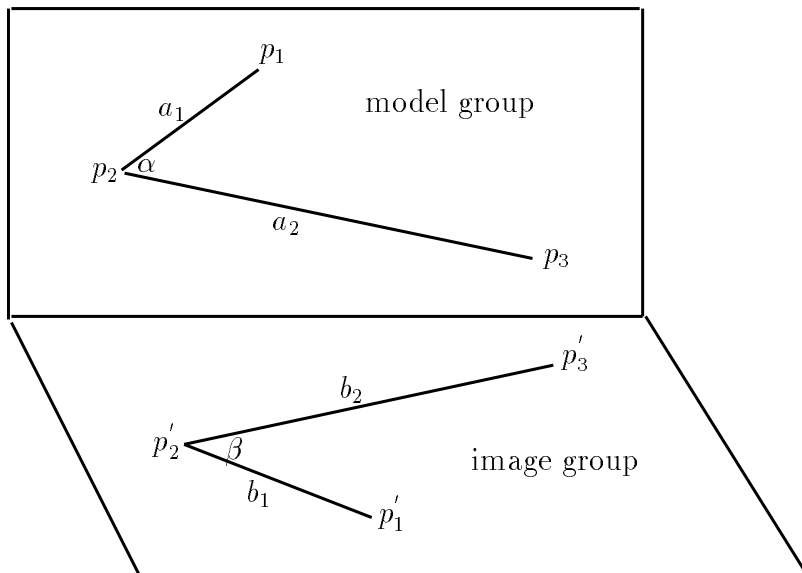


Figure 1: Projected Model Group

$$z = \log(\beta/\alpha) \qquad t = \log\left(\frac{b_1 a_2}{b_2 a_1}\right)$$

$$u = (z^2 + t^2)^{1/2}$$

and the coefficients are given by:

$$\begin{array}{llll} x_1 = 1.9221 & x_2 = 66.811 & x_3 = 0.017544 & x_4 = 1.3850 \\ x_5 = 3.9863 & x_6 = 44.5797 & x_7 = 0.31867 & x_8 = 8.4830 \\ x_9 = 0.48451 & x_{10} = 6.05551 & x_{11} = 2.8818 & x_{12} = 21.513 \end{array}$$

Ben-Arie presents two recognition schemes using the probabilistic peaking effect. The first uses an A^* search technique [Pearl, 1984] as in Ben-Arie and Meiri [1987] but uses as a cost function the result of a stochastic labeling algorithm based on relaxation [Rosenfeld *et al.*, 1976] that uses the probabilistic peaking effect. For objects with 12 features the stochastic labeling algorithm labeled 91% of the features correctly, with worse results for larger objects. No extraneous features were considered in these experiments. The second algorithm matches angular primitives based on which has the largest bayesian probability of matching. The angular primitives of the algorithm consist of an angle and the ratio of lengths of two arms. Three points are necessary to determine this information. So, unless other information is used, $O(m^3)$ model features and $O(n^3)$ image features exist. Note that any additional information used would be of equal value in the alignment method.

In addition, each feature is labeled correctly only 53% of the time. The experiments on synthetic data were performed by projecting the model points using the orthographic projection and without noise or extraneous points.

We use the probabilistic peaking effect to eliminate a match in the alignment method if the probability density for the match is below an empirically determined constant.

4 Eliminating Matches Using the Condition Number

Section 2 gave the solution for the two-dimensional affine transformation as follows:

$$L = \begin{bmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{bmatrix}$$

$$Lb'_m = b'_i \qquad Lc'_m = c'_i$$

These equations can be transformed into:

$$ML_1 = I_1 \qquad ML_2 = I_2$$

where

$$M = \begin{bmatrix} b'_{m_x} & b'_{m_y} \\ c'_{m_x} & c'_{m_y} \end{bmatrix}$$

$$L_1 = \begin{bmatrix} l_{11} \\ l_{12} \end{bmatrix} \qquad L_2 = \begin{bmatrix} l_{21} \\ l_{22} \end{bmatrix}$$

$$I_1 = \begin{bmatrix} b'_{i_x} \\ c'_{i_x} \end{bmatrix} \qquad I_2 = \begin{bmatrix} b'_{i_y} \\ c'_{i_y} \end{bmatrix}$$

Our localization of the image points will have some error, of course. Therefore, our solutions for L_1 and L_2 will also have some error. Let \mathcal{I}_1 , \mathcal{I}_2 , \mathcal{L}_1 , and \mathcal{L}_2 denote the correct values of I_1 , I_2 , L_1 , and L_2 , and let δI_1 , δI_2 , δL_1 , and δL_2 denote their errors such that:

$$\mathcal{I}_1 = I_1 + \delta I_1 \qquad \mathcal{I}_2 = I_2 + \delta I_2$$

$$\mathcal{L}_1 = L_1 + \delta L_1 \qquad \mathcal{L}_2 = L_2 + \delta L_2$$

From basic matrix computations we can bound the error on δL_1 and δL_2 as follows [Watkins, 1991]:

$$\frac{\|\delta L_1\|}{\|L_1\|} \leq \kappa(M) \frac{\|\delta I_1\|}{\|I_1\|} \qquad \frac{\|\delta L_2\|}{\|L_2\|} \leq \kappa(M) \frac{\|\delta I_2\|}{\|I_2\|}$$

where $\|\cdot\|$ is any vector norm (and its induced matrix norm) and $\kappa(M) = \|M\| \cdot \|M^{-1}\|$ is the condition number of M . At present we use the Frobenius norm to determine the condition number even though it is not an induced norm because we have found it to eliminate matches correctly more often than other norms.

So, if M has a large condition number we may have large errors δL_1 and δL_2 . The bounds given are tight. For some direction of δI_1 and δI_2 we have equality. For some directions of δI_1 and δI_2 they are far less accurate. Still it is more likely that a model group will result in accurate values of L_1 and L_2 if the group has a small condition number, as the experiments we describe in Section 7 show.

5 Eliminating Matches Using the Norm of Image Points

From the error bounds in the previous section, we see that if $\frac{\|\delta I_1\|}{\|I_1\|}$ or $\frac{\|\delta I_2\|}{\|I_2\|}$ is large, we may have a large relative error in L_1 or L_2 . Specifically, small values of $\|I_1\|$ and $\|I_2\|$ may produce large errors. Actually, neither of these alone is enough to produce large effect, as we will show here. The equations to solve for L_1 and L_2 from section 4 may be transformed into:

$$\begin{bmatrix} M & 0 \\ 0 & M \end{bmatrix} \begin{bmatrix} L_1 \\ L_2 \end{bmatrix} = \begin{bmatrix} I_1 \\ I_2 \end{bmatrix}$$

Small values of $\|I_1\|$ or $\|I_2\|$ can result in large relative error of L_1 or L_2 , but this is partially a consequence of $\|L_1\|$ or $\|L_2\|$ being small. If one of these is small compared to the other (e.g. $\|L_1\| \ll \|L_2\|$) a large relative error in the small one will not be significant in the computation of $sR^+(p_m - p_{m_0})$. So, we see it is only when both $\|I_1\|$ and $\|I_2\|$ are both small that we may experience problems. Using this analysis we eliminate an image group k if:

$$\left\| \begin{bmatrix} I_{1_k} \\ I_{2_k} \end{bmatrix} \right\|_2 < K \max_i \left\| \begin{bmatrix} I_{1_i} \\ I_{2_i} \end{bmatrix} \right\|_2$$

where I_{1_i} refers to the image values for the i th image group and $K \leq 1.0$ is an empirically determined constant.

6 Eliminating Matches Using the Model Group Area

Even when the model group produces a condition number of one (the best case), it is possible that the group leads to a large error in the computation of the transformation matrix, as can be seen from the following analysis, similar to that done for the affine coordinates of Lamdan *et al.* [1988] in Costa *et al.* [1989]. Assuming non-singularity we have:

$$\begin{bmatrix} l_{11} \\ l_{12} \end{bmatrix} = \begin{bmatrix} b'_{m_x} & b'_{m_y} \\ c'_{m_x} & c'_{m_y} \end{bmatrix}^{-1} \begin{bmatrix} b'_{i_x} \\ c'_{i_x} \end{bmatrix}$$

Computing this inverse we get:

$$l_{11} = \frac{c'_{m_y} b'_{i_x} - b'_{m_y} c'_{i_x}}{b'_{m_x} c'_{m_y} - b'_{m_y} c'_{m_x}}$$

Substituting the original model and image values, this becomes:

$$l_{11} = \frac{(c_{m_y} - a_{m_y})(b_{i_x} - a_{i_x}) - (b_{m_y} - a_{m_y})(c_{i_x} - a_{i_x})}{(b_{m_x} - a_{m_x})(c_{m_y} - a_{m_y}) - (b_{m_y} - a_{m_y})(c_{m_x} - a_{m_x})}$$

Similarly, l_{12} , l_{21} , and l_{22} are:

$$l_{12} = \frac{(c_{m_x} - a_{m_x})(b_{i_x} - a_{i_x}) - (b_{m_x} - a_{m_x})(c_{i_x} - a_{i_x})}{(b_{m_x} - a_{m_x})(c_{m_y} - a_{m_y}) - (b_{m_y} - a_{m_y})(c_{m_x} - a_{m_x})}$$

$$l_{21} = \frac{(c_{m_y} - a_{m_y})(b_{i_y} - a_{i_y}) - (b_{m_y} - a_{m_y})(c_{i_y} - a_{i_y})}{(b_{m_x} - a_{m_x})(c_{m_y} - a_{m_y}) - (b_{m_y} - a_{m_y})(c_{m_x} - a_{m_x})}$$

$$l_{22} = \frac{(c_{m_x} - a_{m_x})(b_{i_y} - a_{i_y}) - (b_{m_x} - a_{m_x})(c_{i_y} - a_{i_y})}{(b_{m_x} - a_{m_x})(c_{m_y} - a_{m_y}) - (b_{m_y} - a_{m_y})(c_{m_x} - a_{m_x})}$$

The area of the triangle formed by a_m , b_m , and c_m is given [Thomas and Finney, 1984] by

$$\text{area}(\triangle a_m b_m c_m) = \pm \frac{1}{2} \begin{vmatrix} a_{m_x} & a_{m_y} & 1 \\ b_{m_x} & b_{m_y} & 1 \\ c_{m_x} & c_{m_y} & 1 \end{vmatrix}$$

which is half the absolute value of the denominator of each of the above equations. Interestingly, the size of the numerator is twice the area of a triangle in a plane where one axis is in model coordinates and one axis is in image coordinates. For l_{11} , the triangle points for the numerator are:

$$p_1 = \begin{bmatrix} a_{i_x} \\ a_{m_y} \end{bmatrix} \quad p_2 = \begin{bmatrix} b_{i_x} \\ b_{m_y} \end{bmatrix} \quad p_3 = \begin{bmatrix} c_{i_x} \\ c_{m_y} \end{bmatrix}$$

$$l_{11} = \pm \frac{\text{area}(\Delta p_1 p_2 p_3)}{\text{area}(\Delta a_m b_m c_m)}$$

If the denominator area is small, this will have the effect of magnifying the errors present in the numerator and the denominator. So, we eliminate a model group k if it satisfies

$$\text{area}(\Delta a_{m_k} b_{m_k} c_{m_k}) < K \max_i \text{area}(\Delta a_{m_i} b_{m_i} c_{m_i})$$

where a_{m_i} refers to the first point of the i th model group and $K \leq 1.0$ is an empirically determined constant.

7 Experimental Results

The techniques presented above have been tested on both randomly generated and real data. This section presents those experiments and results.

7.1 Random Data

To determine the efficacy of the affine transformation as an approximation to the perspective transformation and to determine the best parameters for eliminating unlikely matches, tests were carried out on random point sets. All of the experiments were conducted on sets of ten points with x , y , and z coordinates randomly distributed between -100 and 100. For each point set a random transformation was generated as follows:

1. Three random angles, uniformly distributed between 0 and 360 were generated and the points were rotated about the x , y and z axes by these angles.
2. Three random displacements, uniformly distributed between -25 and 25 were generated, and the points were displaced by these values.
3. The points were projected by the full perspective projection onto the image plane using subpixel accuracy. Various object distances were used to test the efficacy of the affine approximation to the perspective projection. In these experiments the focal length was varied with the distance to keep the size of the images approximately the same.

4. Gaussian noise was added to the each image point's x and y coordinates. Various standard deviations were used in the tests of the affine approximation to the perspective project. A standard deviation of $\sigma = 2.0$ was used for the remaining tests.

Each experiment counted the number of transformations for which the closest transformed model point to each image point is the correct match. Thus, a transformation was only considered successful if each corresponding model and image point were brought close together. For these experiments each model point was projected onto the image and no extraneous image points were included. Since we examine every possible group of three image points, and each has a model group in actual correspondence with it, there are

$$\sum_{i=1}^{10} \sum_{j=i+1}^{10} \sum_{k=j+1}^{10} 1 = 120$$

actual correspondences examined for each random object. Each experiment was executed on 100 random objects.

Figure 2 shows the percentage of actual correspondences which produced successful transformations using objects at varying ratios of object distance to object depth. The three plots are for images with no noise, images with gaussian noise of standard deviation 2.0 added to the image feature coordinates, and images with gaussian noise of standard deviation 5.0 added to the image feature coordinates. The x -axis is the ratio of object distance to approximate depth of the object. We see that for each of the three plots serious degradation of the percentage correct begins at approximately a ratio of 8, which we conclude is the maximum ratio for which the affine approximation to the perspective projection is accurate for use with the alignment method. Since there are usually many correct image groups, the alignment method can find correct transformations even if this ratio is less than 8.

For the remainder of the experiments, we use gaussian noise with standard deviation 2.0, and objects at a distance such that the ratio of object distance to object depth is 10. Figure 3 shows the percentage of total matches examined (not eliminated) and the percentage of correct matches examined for various values of each of the elimination parameters. The most powerful parameter is the peaking probability, as seen by the large distance between the plots. The condition number and model group area parameters appear to be useful for elimination as well, while the image group norm appears to be of little use.

Figure 4 displays the ratio of incorrect matches examined to correct matches examined for various values of each of the elimination parameters. The total remaining ratio and the marginal change from the previous point are shown. Each of the parameters appears to have a graceful degradation, except for the norm of image points (Figure 3.d), which is also the only parameter where the marginal ratio falls below the remaining ratio (indicating that the total remaining ratio rose for that data point.)

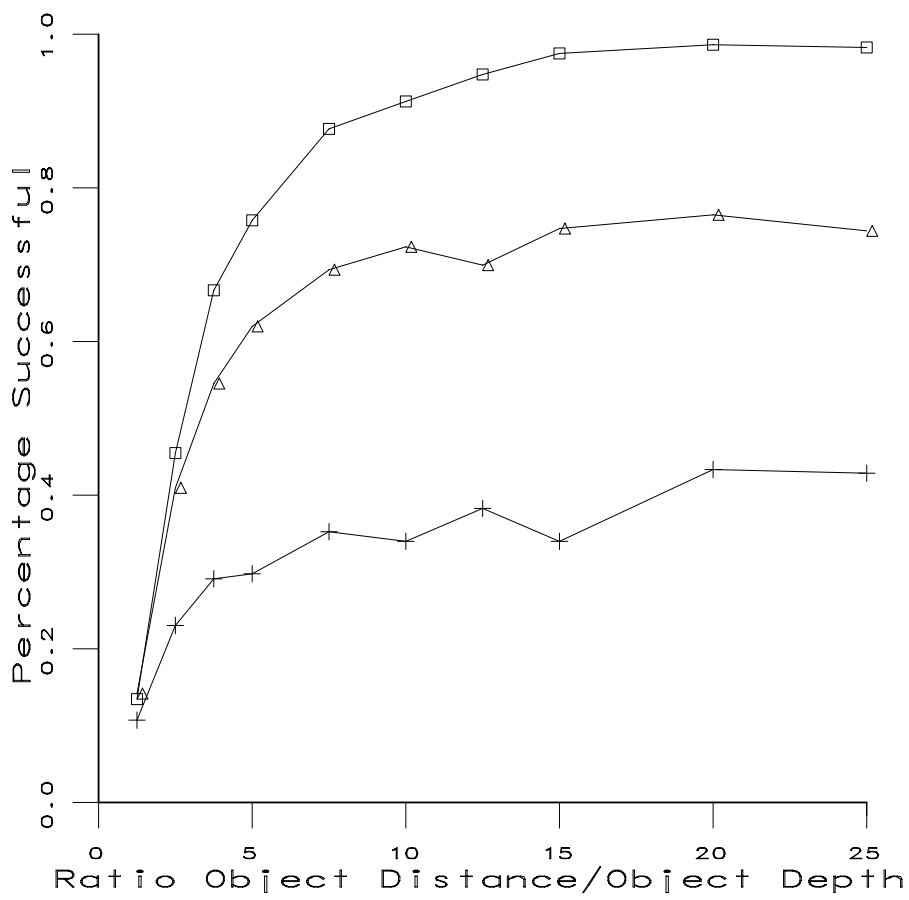


Figure 2: Percentage of successful transformation with various object distances. \square : no noise, \triangle : noise ($\sigma = 2.0$), $+$: noise ($\sigma = 5.0$)

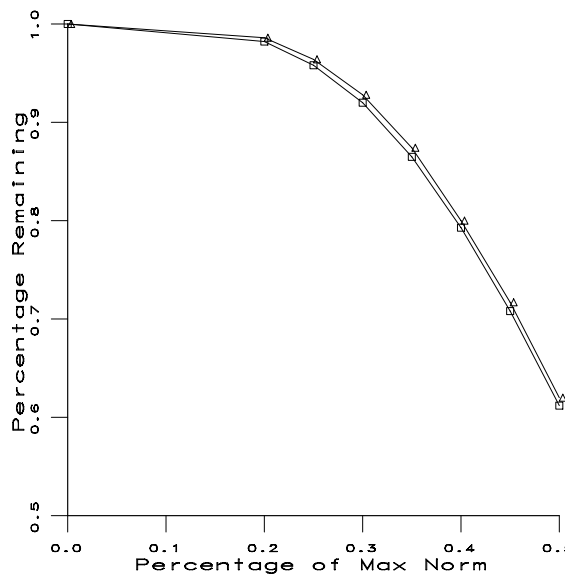
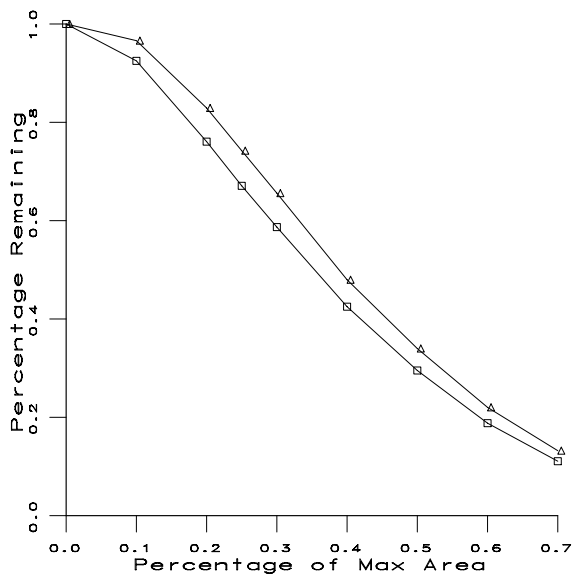
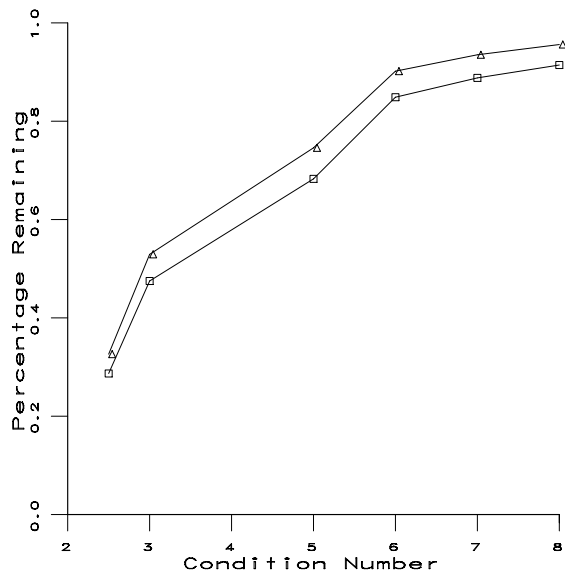
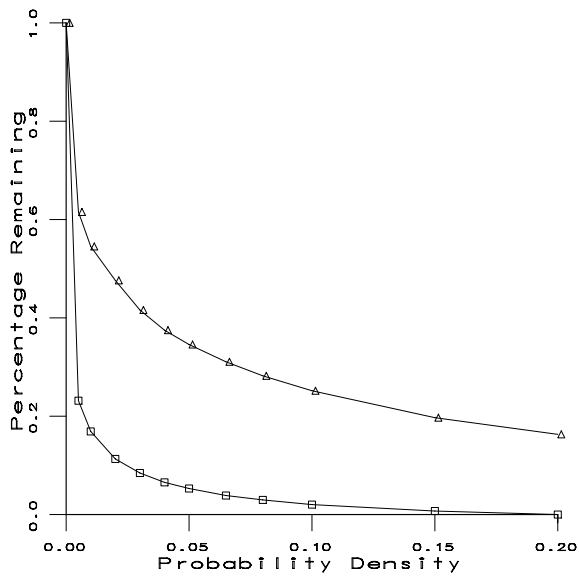


Figure 3: Percentage of total and correct matches examined for various values of each elimination parameter. \square : total, \triangle : correct.

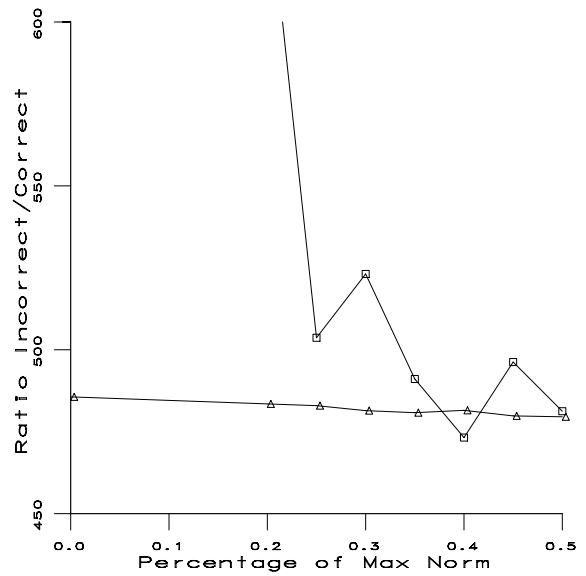
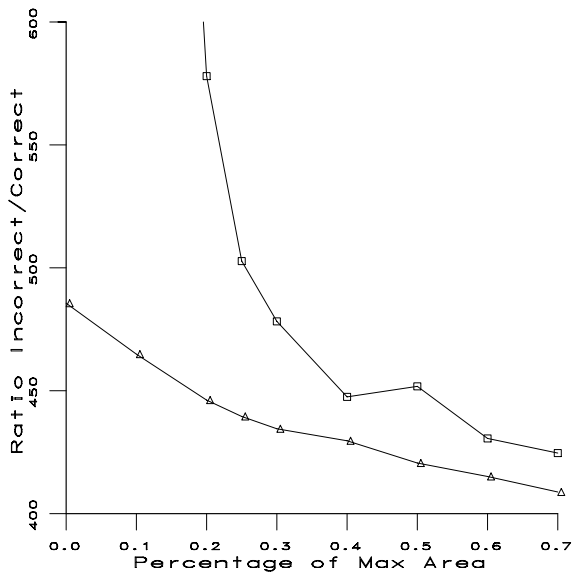
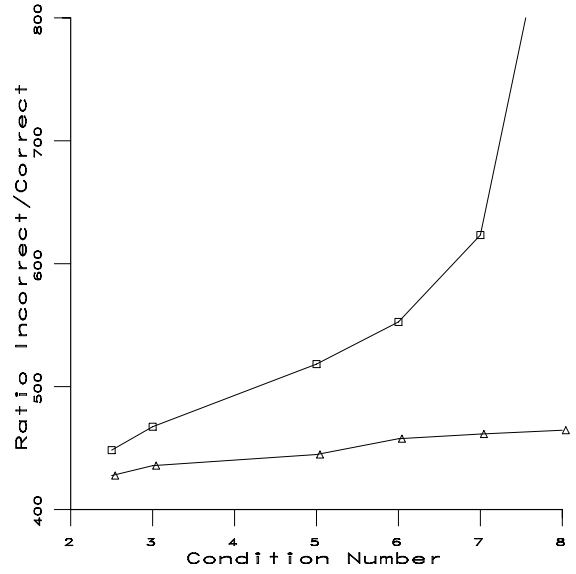
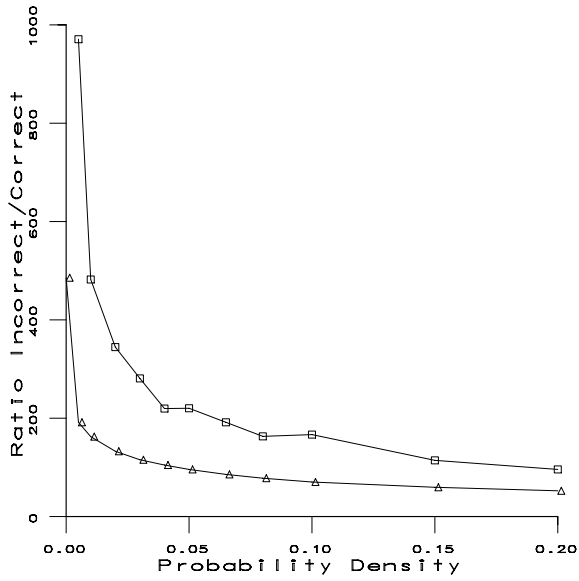


Figure 4: Marginal and remaining ratios of incorrect matches to correct matches for various values of each elimination parameter. \square : marginal, \triangle : remaining.

Using all of the above parameters at reasonable values (peaking probability: 0.08, condition number: 6.0, percentage of max area: 0.3, percentage of max norm: 0.3), 3.1% of all possible matches were examined while 17.9% of correct matches were examined. So, the algorithm examined less than 1/30 of the total matches, yet still examined more than enough correct matches to produce a correct transformation.

7.2 Real Images

The techniques were tested on real images (see Figure 5.a for an example.) The procedure to recognize the model was as follows:

1. Selected model points were measured by hand.
2. The Canny edges [Canny, 1986] were found.
3. Corners were selected by hand.
4. For each group not eliminated, the alignment transformation was found.
5. A quick verification routine determined if the rest of model points were transformed close to image points.
6. If the quick verification routine scored high enough, a more detailed verification routine was executed. This routine determined if edges present in the model were also present in the image.
7. For the best scoring transformation additional matches were determined and the least-squared error affine transformation was determined using a method similar to that of Stein and Medioni [1992].

Figure 5.a shows an image of a stapler that was used to test the elimination techniques. Figure 5.b shows the Canny edges of this image, and 5.c shows the model points as transformed by the highest scoring transformation with some edges drawn in. Figure 5.d is the transformed model overlaid on the Canny edges. For this image, the algorithm using elimination examined 1.8% of the total matches and 37.0% of the correct matches (compared with 3.1% and 17.9% for the random points.) The explanation for the better performance for the real image points sets lies in the fact that the model groups that would appear in unlikely positions (i.e. such that they are highly foreshortened) very often have one or more points occluded by the object itself, while our experiments on random point sets assume no self-occlusion. We therefore expect results on observed points in real images to be consistently better than the experimental results from our random data.

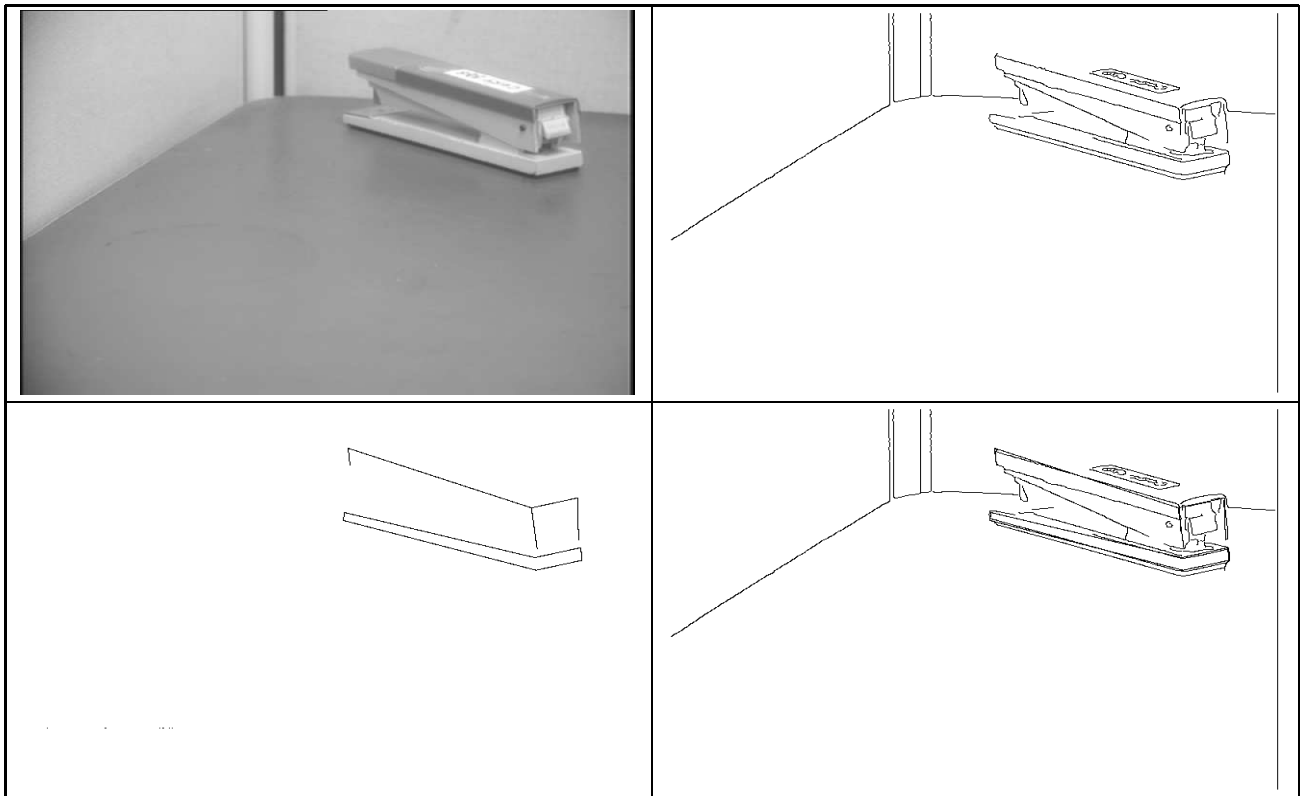


Figure 5: (a) Input image (b) Canny edges (c) Recognized model with edges drawn in (d) Recognized model overlaid on Canny edges

8 Discussion

We will examine the speedup and probability of a false negative produced by these techniques under two recognition models:

1. Possible matches between model groups and image groups are enumerated (either combinatorically or by some grouping process, it does not matter.) Each match that is not eliminated receives a score from some verification process. The best scoring match is accepted if the score meets some criteria.
2. Possible matches are generated in some order (again, either combinatorically or by a grouping process.) Matches are examined in this order. Each examined match that is not eliminated receives a score from the verification process. If this score meets some criteria it is accepted and the remainder of the matches are not examined.

The speedup will be defined as the expected number matches that must be verified by algorithm without using the elimination techniques divided by the expected number when using the techniques. We do not consider the overhead necessary to determine if a match is eliminated in the speedup since this process is $O(1)$ per match and the verification step must be least $O(m \log n)$. Let n be the total number of matches examined, p be the percentage of total matches not eliminated, ν be the total number of matches examined that produce a correct transformation, and ρ be the percentage of these matches not eliminated.

In the first model, we examine np matches when using these techniques and n matches when not using these techniques, so the speedup is simply $\frac{1}{p}$. In the second model, if we assume that there is an equal chance of each possible match being chosen, we have a hypergeometric distribution. For large values of n , this can be approximated by the binomial distribution. The expected number of matches that must be verified when not using the elimination techniques is then approximately $\frac{n}{\nu}$. When using the elimination techniques the expected number of matches that must be verified is approximately $\frac{pn}{\rho\nu}$. So, we see that the expected speedup is about $\frac{\rho}{p}$. This analysis assumes that $\nu > 0$, that is, that a correct match exists to be found. If $\nu = 0$ either because the model is not present in the image or because none of the matches in actual correspondence produce a good enough transformation, then the speedup is the same as for model 1 ($\frac{1}{p}$). Table 1 shows the expected speedup for some values of the elimination parameters. Impressive speedups are attained for recognition model 1 and for model 2 when $\nu = 0$. The speedups for recognition model 2 when $\nu > 0$ are more modest. We note that the case $\nu = 0$ is the common case, since each model must be examined and there are usually few of the models present in the image.

Assuming that the probability of a correct match being eliminated is independent of whether other correct matches have been eliminated (this assumption will be discussed below), the probability of a false negative as a result of eliminating correct matches is $(1 - \rho)^\nu$ for both models. Table 2 shows the probability of a false negative resulting from eliminated matches for the values of ρ in

peaking prob.	condition number	model area	image norm	ρ	p	$\frac{1}{\bar{p}}$	$\frac{\rho}{\bar{p}}$
0.20	4	0.5	0.5	.054	.006	179.81	9.71
0.15	5	0.4	0.4	.092	.012	83.47	7.70
0.10	6	0.3	0.3	.118	.018	55.24	6.52
0.05	7	0.2	0.2	.279	.058	17.17	4.79

Table 1: Speedups for various elimination parameters

ρ	$\nu = 25$	$\nu = 50$	$\nu = 75$	$\nu = 100$
.054	.250	.062	.016	3.88×10^{-3}
.092	.090	8.02×10^{-3}	7.18×10^{-4}	6.44×10^{-5}
.118	.043	1.88×10^{-3}	8.13×10^{-5}	3.52×10^{-6}
.279	2.81×10^{-4}	7.88×10^{-8}	2.21×10^{-11}	6.22×10^{-15}

Table 2: Probability of a false negative for various values of ρ and ν .

Table 1 and various values of ν . As might be expected, for small values of ν there is a non-negligible probability of our techniques causing a false negative. But, we expect that a large number of correct matches will be found since, if even 10 points from a model are present in an image, there are 120 model groups. Grouping algorithms should be able to find a significant number of these model groups. For reasonable values of ν , the probability of a false negative from eliminating correct matches becomes negligible. We mention again that we expect the value of ρ to be larger for real images than for our random model groups, so the probability of false negatives for real images is expected to be much lower than those calculated here.

We now address the question of the independence of the probability of correct matches being eliminated. More specifically, we want to know if it is possible for some object to be in an orientation for which all model groups appear in unlikely configurations in an image. A model group appears in an unlikely configuration when it is viewed from a position with some proximity to the plane the points in the group lie in. Objects that are not nearly flat cannot have the viewing direction nearly coplanar with each model group, so these objects are not a problem. A nearly flat object rotated such that it is very foreshortened in the image may produce angles and/or distance ratios far from the probability peaks. Such images would not benefit much from the elimination of matches based on the peaking parameter, but problems with such images are common to most object recognition systems including the human visual system. Two methods that could help alleviate this problem are to continue relaxing the parameters until most matches have been examined or using special case techniques for recognizing flat [Lamdan *et al.*, 1988] or nearly-flat objects [Arbter *et al.*, 1990].

9 Conclusions

We have presented techniques that greatly reduce the number of matches that must be examined in the alignment method through use of the probabilistic peaking effect and error criteria, thus greatly increasing the speed at which objects can be recognized. Experimental results were given that showed that these techniques work in practice and still result in a correct transformation being found.

Acknowledgements

The author would like to thank 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 and NSF Presidential Young Investigator Grant IRI-8957274 to Jitendra Malik.

References

- [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.
- [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.
- [Canny, 1986] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–697, November 1986.

- [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.
- [Costa *et al.*, 1989] M. Costa, R. Haralick, T. I. Phillips, and L. Shapiro. Optimal affine-invariant point matching. In *SPIE Conference on Artificial Intelligence and Applications*, pages 515–530, March 1989.
- [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.
- [Huttenlocher and Ullman, 1987] D. P. Huttenlocher and S. Ullman. Object recognition using alignment. In *Proceedings of the International Conference on Computer Vision*, pages 102–111, 1987.
- [Huttenlocher and Ullman, 1988] D. P. Huttenlocher and S. Ullman. Recognizing solid objects by alignment. In *Proceedings of the DARPA Image Understanding Workshop*, pages 1114–1124, 1988.
- [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.
- [Lamdan *et al.*, 1988] Y. Lamdan, J. T. Scharz, 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.
- [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.
- [Mundy *et al.*, 1988] J. L. Mundy, A. J. Heller, and D. W. Thompson. The concept of an effective viewpoint. In *Proceedings of the DARPA Image Understanding Workshop*, pages 651–659, 1988.
- [Pearl, 1984] J. Pearl. *Heuristics: Intelligent Search Strategies for Computer Problem Solving*. Addison-Wesley, 1984.
- [Rosenfeld *et al.*, 1976] A. Rosenfeld, R. A. Hummel, and S. W. Zucker. Scene labeling by relaxation operations. *IEEE Transactions on Systems, Man, and Cybernetics*, 6(6):420–433, June 1976.
- [Stein and Medioni, 1992] F. Stein and G. Medioni. Structural indexing: Efficient 3-d object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):125–145, February 1992.

- [Thomas and Finney, 1984] G. B. Thomas and R. L. Finney. *Calculus and Analytic Geometry*. Addison-Wesley, sixth edition, 1984.
- [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.
- [Watkins, 1991] D. S. Watkins. *Fundamentals of Matrix Computations*. John Wiley and Sons, 1991.