Robust Feature Selection for Object Recognition using Uncertain 2D Image Data*

Tarak L. Gandhi
gandhi@cse.psu.edu
Dept. of Computer Science and Engineering
The Pennsylvania State University
University Park, PA 16802

Octavia I. Camps
camps@whale.ece.psu.edu
Dept. of Electrical Engineering
The Pennsylvania State University
University Park, PA 16802

Abstract

The use of a *small* set of features is recurrent in the object recognition literature. If the image data is perfect with no sensor uncertainty and there are not incorrect feature correspondences between the model and the image, then the pose of the object can be computed with no error using these *few* correspondences. However, in most real cases the noise in the data will propagate into the pose. Moreover, the extent of the effect of the uncertainty will depend on the selection of the correspondences used to compute it. In this paper we address the problem of how to select these correspondence so that the effect of the data uncertainty on the pose estimation is minimized.

1 Introduction

Most model-based computer vision systems attempt to recognize and locate 3D objects from a 2D image of a scene by pairing features from a set of stored models with features extracted from the image. These correspondences are found using techniques such as interpretation trees [10, 7, 2], hashing [27, 5, 9, 26], alignment [17], bipartite search [20], and automated programming [1]. The pairings are such that the features in the image can be obtained (approximately) by applying a geometric transformation to their corresponding model features. This transformation is usually referred as the *pose* of the object, that is the position of the object with respect to a coordinate system. Most methods to compute the pose use a *few* point-to-point [14, 6, 12, 11] or line-to-line [21, 22] correspondences. If the data is perfect with no sensor uncertainty and with no incorrect correspondences, then the pose is exact, and the transformed model features exactly coincide with the image features. However, in most real cases the noise in the data will propagate into the pose. Moreover, the extent of the effect of the uncertainty depends on the correspondences used to compute it.

In particular, an important issue that affects the performance of a recognition system is the fact that the accuracy of a pose computed using a small number of correspondences can be very different depending on which correspondences are selected, even when the same number of correspondences is used.

Recently, Grimson *et al* [12] presented a detailed study of how sensor uncertainty in the data propa-

gates into the pose when it is computed using three point correspondences and the method given in [11]. Furthermore, they used their results to analyze the effects of sensor noise in the performance of systems that use feature alignment or hashing schemes to do object identification. However, at the present time there are no methodologies available to improve the performance of a recognition system by incorporating studies such as [12] into the selection process of the feature correspondences. Thus, designers are forced to build recognition systems in an iterative fashion, trying different feature selection heuristics until the desired level of performance is achieved.

2 Statement of the Problem

The use of a small set of features is recurrent in the literature. Perceptual groupings was first suggested by Lowe [23]. Henikoff and Shapiro [15] defined *interesting patterns* formed by triplets of line segments and found that they were useful in reducing the number of hypothesized models. In the work by Mohan and Nevatia [24], the systems 3DPO [16] and 3D-POLY [25] a few "local" or "kernel" set of features were used. Hansen [13] used a set of filters in order to reduce the number of features to be considered. Flynn [8] proposed the use of an utility measure of the features in order to reduce the number of hypothesis made by a hashing scheme.

Ikeuchi and Kanade [18], Chen and Mulgaonkar [4], and Camps [3] used the concept of feature detectability to rank features in decreasing order of detectability.

In spite of all this research activity, the selection of "good" features to be matched in object recognition remains difficult. In this paper we address the following problem:

> Let N be the number of model points and $n \leq N$ be the number of points that are actually used to compute the pose of the object. Then, find the subset of n model points such that the effect of the data uncertainty in the estimation of the pose is minimized.

3 Definitions and Notation

In photogrammetric terminology, the *exterior orientation* of a camera is specified by all the parameters (three rotation angles and a translation vector) that

determine the pose of the camera with respect to the world reference frame.

The relation between the camera and the world reference frames is given by a rotation and a translation. A point m represented in the world reference frame by the coordinates $(X\ Y\ Z)'$ is represented by the coordinates $(x\ y\ z)'$ in the camera reference frame where

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = R \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + d \ , \qquad (1)$$

R is a 3×3 rotation matrix, and d is a 3×1 translation vector.

Once the point m is represented in the camera reference frame, it can be projected using the perspective projection:

$$P_{(R,d)}(m) = \begin{pmatrix} u \\ v \end{pmatrix} = \frac{f}{z} \begin{pmatrix} x \\ y \end{pmatrix} \qquad (2)$$

where $P_{(R,d)}(m)$ denotes the perspective projection of m and f is the focal length of the camera.

In the sequel we represent a 3D model object by a $3\times N$ matrix M,

$$M = \begin{pmatrix} m_1 & m_2 & \dots & m_N \end{pmatrix} \qquad (3)$$

where the column m_i is given by the world coordinates of the model point i. The perspective projection of the model M is denoted by $P_{(R,d)}(M)$ and is represented by a $2N \times 1$ matrix C_o,

$$P_{(R,d)}(M) = C_o = \begin{pmatrix} c_{o1} \\ c_{o2} \\ \vdots \\ c_{oN} \end{pmatrix} \qquad (4)$$

where c_{oi} is given by the image coordinates of the perspective projection of the model point i. Finally, the *observed* image points assigned to the object points by some mapping algorithm are represented by $2N \times 1$ matrix C,

$$C = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{pmatrix} \qquad (5)$$

where c_i is given by the image coordinates of the point corresponding to the model point i. If all the correspondences are correct, then C_o and C are related by

$$C = C_o + \Delta C_o \qquad (6)$$

where ΔC_o is additive mean zero Gaussian random noise having covariance matrix Σ.

4 Pose Estimation

In this paper we will limit ourselves to the case where the camera is located on a sphere of radius r centered at the world reference frame origin, and such that the translation vector is known. The effect of these constraints is that the number of unknown exterior parameters is reduced from six to three, since the translation d is known. Given enough pairs of corresponding 2D and 3D points and an initial approximate solution, the rotation R can be found by using a non-linear least-square technique like the one described in [14, 23].

Let M be the set of model points, C be the set of observed image points, Σ be the noise covariance matrix,

and d be the translation vector. Then, the rotation matrix R can be found by minimizing the least-squares criterion

$$\epsilon^2 = \left(C - P_{(R,d)}(M)\right)' \Sigma^{-1} \left(C - P_{(R,d)}(M)\right) \ . \quad (7)$$

Assuming that $\Sigma = \sigma^2 I$, minimizing ϵ^2 is equivalent to minimize

$$\sum_i \left\| c_i - P_{(R,d)}(m_i) \right\|_2^2 = \left\| C - P_{(R,d)}(M) \right\|_F^2 \quad (8)$$

where $\|.\|_2$ and $\|.\|_F$ denote the Euclidean and the Frobenius norms respectively.

Since $P_{(R,d)}$ is a non-linear operator, an iterative procedure must be used to solve for R. The rotation matrix R can be found by linearizing $P_{(R,d)}$ around an approximate solution, solving for small corrections and adjusting the solution iteratively.

Let $r_x^{(k)}$, $r_y^{(k)}$, and $r_z^{(k)}$ represent the world reference axes rotated by the rotation matrix $R^{(k)}$ for the kth iteration:

$$R^{(k)} = \begin{pmatrix} r_x^{(k)'} \\ r_y^{(k)'} \\ r_z^{(k)'} \end{pmatrix} \ . \qquad (9)$$

The rotation $R^{(k)}$ can be adjusted in a fast and easy way by premultiplying it with three correction-rotation matrices:

$$R^{(k+1)} = R_{h_z}^{(k)} R_{h_y}^{(k)} R_{h_x}^{(k)} R^{(k)} \qquad (10)$$

where $R_{h_x}^{(k)}$, $R_{h_y}^{(k)}$, and $R_{h_z}^{(k)}$ are rotations around the axes $r_x^{(k)}$, $r_y^{(k)}$, and $r_z^{(k)}$ with small angles $h_x^{(k)}$, $h_y^{(k)}$, and $h_z^{(k)}$ respectively. The optimal correction is found by solving the equation

$$J^{(k)} h^{(k)} = e^{(k)} \qquad (11)$$

where $J^{(k)}$ is the $2N \times 3$ Jacobian matrix,

$$J^{(k)} = \begin{pmatrix} j_1 \\ j_2 \\ \vdots \\ j_N \end{pmatrix}^{(k)} \qquad (12)$$

$$j_i = \frac{\partial c_i}{\partial h} = \frac{1}{z_i} \begin{pmatrix} -u_i q_i & f s_i + u_i p_i & -f q_i \\ -f s_i - v_i q_i & v_i p_i & f p_i \end{pmatrix} \ ,$$

$$\begin{pmatrix} p_i \\ q_i \\ s_i \end{pmatrix}^{(k)} = R^{(k)} m_i = R^{(k)} \begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} \ ,$$

$$\begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix}^{(k)} = R^{(k)} m_i + d = \begin{pmatrix} p_i \\ q_i \\ s_i \end{pmatrix}^{(k)} + d,$$

$$\begin{pmatrix} u_i \\ v_i \end{pmatrix}^{(k)} = P_{(R^{(k)},d)}(m_i) = \frac{f}{z_i^{(k)}} \begin{pmatrix} x_i \\ y_i \end{pmatrix}^{(k)} \ ,$$

$h^{(k)}$ is the vector of the unknown corrections for which we are solving,

$$h^{(k)} = (h_x^{(k)}\ h_y^{(k)}\ h_z^{(k)})' \qquad (13)$$

and $e^{(k)}$ is a 3 × 1 vector with the error measured in the image:

$$e^{(k)} = C - P_{(R^{(k)},d)}(M) . \qquad (14)$$

Thus, the rotation $R^{(k)}$ at iteration k is updated using equation (10) with

$$h^{(k)} = J^\dagger \left(C - P_{(R^{(k)},d)}(M) \right) . \qquad (15)$$

5 Sensitivity Analysis

The sensitivity of the pose estimation algorithm to the noise in the data is given by the amount of perturbation on the rotation matrix due to a small perturbation in the data. In the sequel, we present a sensitivity analysis similar to the one presented in [14].

Let M be the matrix of the model points, C be the matrix of the *observed* image points, R be the true rotation matrix, and $\hat{R}$ be the rotation matrix estimated using the iterative least squares method described in the previous section. Then,

$$P_{(\hat{R},d)}(M) \cong P_{(R,d)}(M) + \nabla_h P_{(R,d)}(M).h \qquad (16)$$

where

$$h = J^\dagger e , \qquad (17)$$

$$e = C - P_{(R,d)}(M) , \qquad (18)$$

and $J^\dagger$ is the pseudo-inverse of J.

The expected value of the squared norm of the correction vector, $||h||^2$, is given by

$$E[h'h] = E[\text{trace}(hh')] . \qquad (19)$$

If e is assumed to be zero-mean Gaussian random noise with covariance matrix $\Sigma = \sigma^2 I$, the expected value of hh' is

$$E[hh'] = E[J^\dagger ee' J^{\dagger'}] = \sigma^2 J^\dagger J^{\dagger'} = \sigma^2 (J'J)^{-1} \qquad (20)$$

and the expected value of $||h||^2$ reduces to

$$E[h'h] = \sigma^2 \text{trace}(J'J)^{-1} . \qquad (21)$$

Thus, the trace of the matrix $(J'J)^{-1}$ can be used as a measure of the sensitivity of the pose to the noise in the data.

6 Selection of Point Correspondences

Let M be the set of N model points, and $n \leq N$ be the number of points that are actually going to be used to compute the pose. In order to minimize the effect of the data uncertainty on the estimated pose, these n points need to be selected carefully. In the previous section we showed that the trace of the matrix $(J'J)^{-1}$ measures the sensitivity of the pose to the noise in the features used in the computation. Thus, the subset of n points that minimizes the trace of the above matrix:

$$\min_{\text{Subsets of } n \text{ points}} \left\{ \text{trace } (J'J)^{-1} \right\} \qquad (22)$$

is a good choice from these considerations. Unfortunately, the number of possible subsets of n points is, in general, too large to attempt to solve this minimization problem directly. However, a suboptimal solution to this problem can be obtained by using an incremental approach:

Given a subset $\mathcal{P}_k \subset M$ with k points, select a point $m \in M$ such that the trace of $(J'J)^{-1}$ is minimized for the extended subset $\mathcal{P}_{k+1} = \mathcal{P}_k \cup \{m\}$.

7 The Greedy Algorithm

In this section we describe an algorithm that can be used by a matching procedure to select the next model feature to be assigned a correspondence. This algorithm is a *greedy* algorithm that finds a model point such that when it is added to the points used so far, the pose computed using the algorithm described in section 4 is robust to the noise present in the data. The algorithm is suboptimal since it has a limited horizon of one point at a time.

7.1 Selection of initial points

The pose estimation algorithm given in section 4 starts with an initial rotation $R^{(0)}$ and then iterates to refine this pose. In order to compute the initial rotation $R^{(0)}$, a minimum of two points correspondences are required. Let m_1 and m_2 be two model points. It can be easily shown that if m_1, m_2, and the origin of the world reference frame are aligned, i.e. $m_2 = \alpha.m_1$, the estimated pose is not unique. Furthermore, if the points m_1 and m_2 are close to the origin, a small perturbation in the coordinates of the corresponding image points leads to a large change in the estimated pose. This suggests the heuristic rule that the initial points should be selected such that the area of the triangle formed by the two model points and the origin is maximum.

7.2 Selection of subsequent points

Once k correspondences have been found the problem of selecting the next correspondence such that the estimated pose is robust reduces to selecting the model point that minimizes the trace of the 3 × 3 matrix $(J'J)^{-1}$.

Consider the set of k model points already assigned to some image point, $\mathcal{P}_k$, the corresponding matrix $J_k'J_k$, and a candidate model point $m \notin \mathcal{P}_k$. The matrix $J_{k+1}'J_{k+1}$ for the extended set $\mathcal{P}_{k+1} = \mathcal{P}_k \cup \{m\}$ can be computed incrementally:

$$J_{k+1}'J_{k+1} = J_k'J_k + j_m'j_m \qquad (23)$$

where j_m is given by equation (13). Thus, the $(k+1)$-th point can be selected by solving for m in

$$\min_m \left\{ \text{trace}(J_k'J_k + j_m'j_m)^{-1} \right\} . \qquad (24)$$

7.3 Handling outliers

The initial rotation $R^{(0)}$ is found by solving a system with four equations and three unknowns such that a least square error criterion is minimized for the two initial model points. If the error of this fit is too high, at least one of the points is likely to be an outlier and a new pair of points is selected.

When a subsequent point is added, one can use the current rotation $R^{(k)}$ to project the model points currently used and compare their location with their corresponding image points. If the distance between these

is higher than a multiple of the standard deviation of the noise, then the point is rejected as an outlier.

If at a given point, too many points are classified as outliers, the initial points are suspected as outliers and the process starts again for a new pair of initial points.

8 Experimental Protocol

The importance of controlled experiments has only recently been stressed in computer vision. Controlled experiments are essential in order to illustrate the validity of a solution presented. We tested the greedy algorithm using artificially generated data as well as real images. In this section we describe the experimental protocol used to test the proposed algorithm, based upon the one presented in [19].

8.1 Data Set Generation

The data for the experiments was generated as follows:
- **Random model generation.** $N_{model} = 50$ random model sets were generated. The model points were generated by choosing points uniformly distributed in a box such that value of each of the coordinates lie in $[-M_{max}, M_{max}]$ where $M_{max} = 5.0cm$.. The number of model points generated in each experiment was $N = 10$.
- **Random exterior parameters generation.** The rotation matrices R_o were generated by selecting random points on a unit sphere and taking the r_z axis as the line from these points to the origin. The r_x and r_y axes lie in the plane perpendicular to the r_z axis. The orientation of r_x axis in that plane was chosen randomly and r_y axis was made perpendicular to it. The translation vector d was kept fixed w.r.t. the rotated coordinate system as $d = (0\ 0\ d_z)'$ where $d_z = 35.0cm$.
- **Image point generation.** The model points were transformed using : $C_o = R_o M + d$. The projected points fall in a box with u,v coordinates within $[-C_{max}, C_{max}]$ where $C_{max} \approx 0.7cm$..

8.2 Noise Modeling

In order to test the algorithm in the presence of noise and outliers the following perturbations were introduced:
- **Rotation perturbation.** The rotation matrix R_o was perturbed by rotating it around its r_x and r_y axes respectively. No rotation was done around r_z axis since this does not result in any change in the sensitivity of the problem. The amount of rotation was varied from 0.05 to 0.3 radians.
- **Image noise generation.** Independent, identically distributed Gaussian noise $N(0, \sigma^2 I)$ was added to the image points to generate points with camera coordinates $C = C_0 + \Delta C_0$. This corresponds to SNR of the order of $\frac{\sigma}{C_{max}}$ in terms of length and its square in terms of area. For each set of model points, $N_{sample} = 20$ such samples were generated.
- **Outlier generation.** The number of outliers was varied as $N_{out} = 1, \ldots, 4$. Camera points were randomly selected and replaced by randomly generated 2D points within the range $[-C_{max}, C_{max}]$.

8.3 Experimental Procedure

The following experiments were conducted:
Experiment I. No noise was added in this experiment. Only the sensitivity of the solution was studied. The greedy method was applied to each of the model sets, varying the number of points used, n, from 2 to N. The obtained greedy sequences, $\mathcal{P}_g$, were compared against the optimal sequence (found by exhaustive search) with the same number of points, $\mathcal{P}_o$. This comparison was done by computing the difference between the traces of the matrix $(J'J)^{-1}$ for each sequence of length 2 to N. Those greedy sequences such that

$$\left| \frac{\text{trace}(J'J)^{-1}}{\text{trace}(J_o'J_o)^{-1}} \right| > Th \qquad (25)$$

where Th is a suitable threshold were rejected. The percentage of rejected sequences was plotted against the length of the sequence n. The value of the trace for the greedy sequences, for sequences that have trace equal to the best trace, and equal to the median of the traces of all possible sequences were plotted. This was done to show that the greedy sequences give better solutions than typical sequences and comparable to the optimal sequence.

Experiment II. This experiment studies the effect of perturbation on the rotation matrix on the greedy sequence. Although the exact pose of the object is unknown, an approximate pose can be obtained at any given point in the algorithm. However, the dependence of the Jacobian matrix on the pose, i.e. second order effects are neglected. Since such perturbation is caused by noise, this experiment indirectly helps in knowing what to expect when noise is present.

For each model set, $N_{sample} = 20$ random perturbations were applied to the original rotation matrix, and new camera points were projected. No further noise nor outliers were added to these points. The greedy solution was computed using the new sets of camera points. Since the rotation matrices were perturbed, the trace values also were perturbed. Comparison of these trace values was made with the optimal value by finding the number of experiments in which the trace found was greater than the trace of the optimal sequence of the same length by more than a threshold value Th, i.e.

$$\left| \frac{\text{trace}(J'J)^{-1}}{\text{trace}(J_o'J_o)^{-1}} \right| > Th \qquad (26)$$

The percentage of rejected sequences was plotted against the length of the sequence n.

Experiment III. This experiment studies the performance of the algorithm in the presence of noise.

For each model set, random Gaussian noise was added to the points. $N_{sample} = 20$ such samples of noise were added for each set. However, no outliers were added. The greedy solution was computed using each of these. Since noise perturbs the pose, the trace values also get perturbed. Comparison of these trace values was made with the optimal sequence of the same length, by finding the number of experiments in which the trace found was greater than 10 % of the trace of

the corresponding optimal sequence. The distance between the actual pose R_o and the computed pose R is defined as the norm of their difference:

$$\|R - R_0\|_2 = \max_{\|x\|_2=1} \|Rx - R_o x\|_2 \qquad (27)$$

This distance corresponds to the largest possible deviation of the coordinates of a unit vector when transformed by R instead of R_o. The theoretical RMS value of this distance is equal to $\sigma_{noise}\sqrt{\text{trace}(J'J)^{-1}}$. In this experiment, a greedy sequence was rejected if

$$\frac{\|R - R_0\|_2}{\sigma_{noise}\sqrt{\text{trace}(J'J)^{-1}}} > Th \qquad (28)$$

For comparison, for each of the samples, a random sequence was used in place of the sequence found by the greedy algorithm. Similar plots were made for these sequences, to show the relative merit of the greedy sequence.

Experiment IV. This experiment studies the performance of the algorithm in the presence of noise and outliers. For each model set, random Gaussian noise was added to the points and outliers were introduced. $N_{sample} = 20$ such samples of noise were generated for each model. The greedy solution was computed using each of these and discarding possible outliers. The traces and the distance between the poses were compared in a similar way to the previous experiment. However, the comparison was done only for the sequences that successfully rejected the outlier points. The percentage of the sequences that failed to detect outliers were also plotted against the length of the sequence n.

9 Results

In this section we discuss the results obtained when the experiments described in the experimental protocol were carried out, as well as the results obtained with real data.

Experiment I The plot of the percentage of models with greedy sequences failing the trace comparison test for several threshold values is given in Fig 1 (a). Fig 1 (b) shows a similar plot using the sequences having median trace. It is observed that the greedy
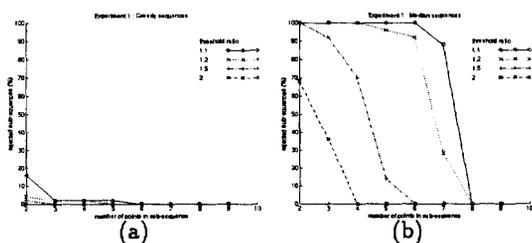
Figure 1: Experiment I. (a) Percentage of greedy sequence failing the trace comparison test. (b) Percentage of median sequences failing the trace comparison test.

sequences fail less than the median sequence, showing the effectiveness of greedy algorithm. The failure rate decreases as the length of the sequence increases. This
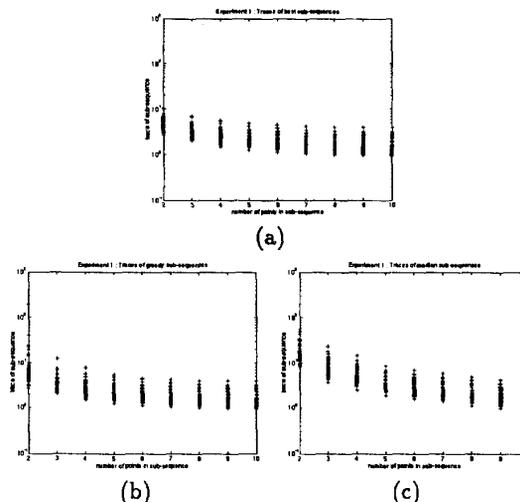
Figure 2: Experiment I. (cont) (a) Trace values for the best sequences. (b) Trace values for the greedy sequences. (c) Trace values for the median sequences.
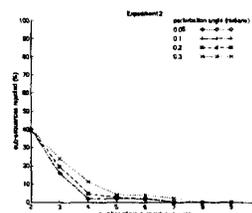
Figure 3: Experiment II. Performance characterization of the greedy algorithm when the rotation is perturbed.

is expected since initially the pose is not known. However, as more points are selected, the greedy criterion takes over, selecting sequences with the best trace at each step. The values of the traces of the best, greedy and median sequences are shown in Fig 2 (a)-(c) respectively.

Experiment II The plot in Fig 3 shows the effect of the perturbation in the rotation on the performance of the greedy algorithm. As the perturbation increases, the chance of failure of the greedy sequence also increases. However, no marked deterioration was observed for perturbations up to 0.1 radians (5.7 degrees).

Experiment III The effect of the data noise on the performance of the algorithm is shown in Fig 4. Fig 4(a) shows the rate of failure using the trace test with threshold ratio of 1.1 and Fig 4(b) the rate of failure using the norm of the rotation difference test with threshold ratio of 1.5. The standard deviation σ_{noise} was varied between 0.01 (SNR of 0.014) and 0.1 (SNR of 0.14). It is observed that the effect on the performance is marked after the σ_{noise} increases beyond 0.05. Similar results for random sequences are shown

in Fig 4(c) and Fig 4(d). It can be easily seen that the number of rejected sequences is much greater for the random sequences, hence highlighting the advantage of the proposed selecting method.
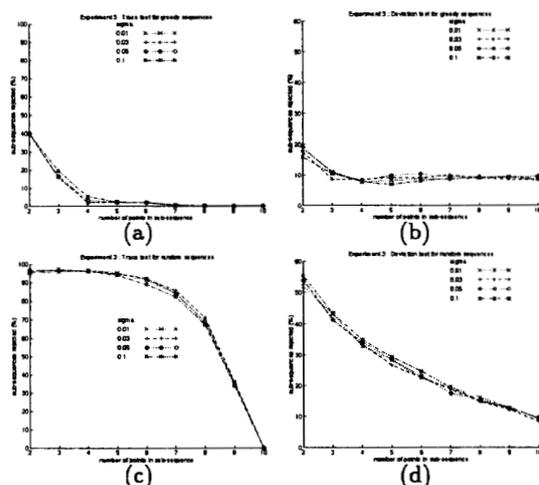
Figure 4: Experiment III. (a) Trace performance for the greedy sequence. (b) Rotation difference performance for the greedy sequence. (c) Trace performance for a set of random sequences. (d) Rotation difference performance for a set of random sequences.

Experiment IV The effect of noise and outliers on the performance of the algorithm is shown in Fig 5. Fig 5(a) shows the rate of failure using the trace test with threshold ratio of 1.1 and Fig 5(b) shows failure rate using norm of difference test with threshold ratio of 1.5. However, these tests were performed only for the cases where the outliers points were properly detected. Fig 5 shows the percentage of sequences where some outlier point was not detected. The noise standard deviation σ_{noise} was fixed at 0.03 and the number of outliers added was varied from 1 to 4. It is seen that the performance deteriorates quite rapidly as the number of outliers is increased. The crossings near the end of some of the curves are accounted by the fact that the number of choices available are lower for sets with more outliers. Hence, the value of the best trace is expected not to be much higher than the trace of the greedy sequences.

10   Real data

The algorithm was also tested on real data. Fig. 6 shows the results obtained with an image of a bookend. Fig. 6(a) shows a grayscale image of the bookend with the model points highlighted. Fig. 6(b) shows the back projection of the model when all the model points are used to compute the pose. Fig. 6(c) shows the back projection of the model onto the image when four random points (circled on the figure) are used to compute the pose. Finally, Fig. 6(d) shows the back projection of the model when four points are selected using the greedy algorithm (circled on the figure). Clearly, the
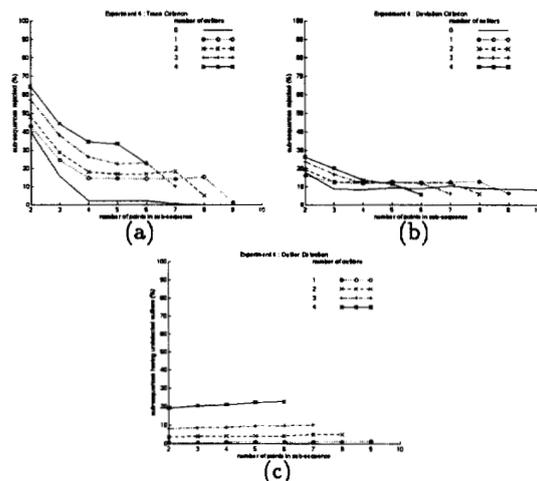
Figure 5: Experiment IV. (a) Trace performance for the greedy sequence. (b) Rotation difference performance for the greedy sequence. (c) Percentage of sequences where some outlier was not detected.

greedy solution is better than the random one, and comparable to the one obtained using all the points.

11   Conclusions

In this paper we studied the noise sensitivity of a point-to-point iterative least square pose estimation procedure when the translation of the camera is constrained. These results were applied in the design of a *robust* algorithm *to select the next feature to be matched* by a recognition system. The proposed algorithm selects the next point such that it minimizes the effect of the data uncertainty on the computation of the pose. Even though the solution is only suboptimal (since the algorithm looks for only a point at a time) our experimental results show that the computed poses using the points provided by the algorithm are comparable with the best possible poses computed with the same number of points.

The algorithm requires the computation of the inverse of $O(N^2)$ 3 ×3 matrices where N is the total number of model points. However, the computation of these matrices is done incrementally, reducing significantly the complexity.

We have laid out a rigorous methodology to characterize the performance of the algorithm in the presence of noise and outliers and we have presented experimental results using both artificial and real data.

Finally, this algorithm can be easily extended to the case where the camera looks in an arbitrary direction, but it remains at a fixed distance from the origin. On the other hand, the most general case when all six exterior parameters are unknown, is more difficult and it is the subject of future research.

References

[1] F. Arman and J. Aggarwal. Automatic Generation of
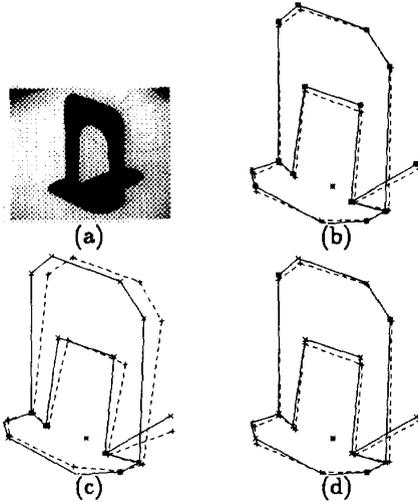
Figure 6: Bookend image. (a) Grayscale image with model points highlighted. (b) Back projection using all the model points. (c) Back projection using four random model points (circled on the figure). (d) Back projection using four points selected using the greedy algorithm (circled on the figure).

Recognition Strategies Using CAD Models. In *Proc. of the IEEE Workshop on Directions in Automated CAD-Based Vision*, pages 124–135, June 1991.

[2] O. I. Camps. *PREMIO: The Use of Prediction in a CAD-Model-Based Vision System*. PhD thesis, Department of Electrical Engineering, University of Washington, Seattle, Washington, 1992.

[3] O. I. Camps, L. G. Shapiro, and R. M. Haralick. Image prediction for computer vision. In J. A.K. and P. Flynn, editors, *Three-dimensional Object Recogntion Systems*. Elsevier Science Publishers BV, 1993.

[4] C. Chen and P. Mulgaonkar. Automatic vision programming. *CVGIP: Image Understanding*, 55(2):170–183, March 1992.

[5] M. Costa, R. Haralick, and L. Shapiro. Optimal affine-invariant point matching. In *Proc. of the International Conference on Pattern Recognition*, pages 233–236, Atlantic City, New Jersey, June 1990.

[6] R. Ellis. Uncertainty estimation for polyhedral object recognition. In *Proc. of the IEEE Conference on Robotic and Automation*, May 1989.

[7] P. Flynn. *CAD-Based Computer Vision: Modeling and Recognition Strategies*. PhD thesis, Michigan State University, 1990.

[8] P. Flynn. Saliencies and Symmetries: Towards 3D Object Recognition from Large Model Databases. In *Proc. of the IEEE Computer Vision and Pattern Recogntion*, pages 322–327, June 1992.

[9] P. Flynn and A. Jain. 3D Object Recognition Using Invariant Feature Indexing of Interpretation Tables. *Computer Vision, Graphics, and Image Processing: Image Understanding*, February 1992.

[10] W. Grimson. The combinatorics of object recognition in cluttered environments using constrained search. In

[11] W. Grimson. *Object Recognition by Computer: The role of geometric constraints*. MIT Press, Cambridge, MA, 1990.

Proc. of the International Conference on Computer Vision, pages 218–227, 1988.

[12] W. Grimson. Recognizing 3D Objects from 2D Images: An Error Analysis. Technical Report A.I. Memo No. 1362, MIT, Cambridge,MA, November 1992.

[13] C. D. Hansen. *CAGD-Based Computer Vision: The Automatic Generation of Recognition Strateegies*. PhD thesis, The University of Utah, 1988.

[14] R. Haralick and L. Shapiro. *Computer and Robot Vision*. Addison-Wesley, 1992.

[15] J. Henikoff and L. Shapiro. Interesting patterns for model-based matching. In *ICCV*, 1990.

[16] P. Horaud and R. Bolles. 3DPO: A system for matching 3-D objects in range data. In A. Pentland, editor, *From Pixels to Predicates*, pages 359–370. Ablex Publishing Corporation, Norwood, New Jersey, 1986.

[17] D. Huttenlocher and S. Ullman. Object recognition using alignment. In *Proceedings of the First International Conference on Computer Vision*, pages 102–111, 1987.

[18] K. Ikeuchi and T. Kanade. Modelling sensor detectability and reliability in the configuration space for model-based vision. Technical Report CMU-CS-87-144, Carnegie-Mellon University, Computer Science Department, July 1987.

[19] T. Kanungo, M. Jaisimha, R. Haralick, and J. Palmer. An experimental methodology for performance characterization of a line detection algorithm. In *SPIE Conference on Optics, Illumination and Image Sensing for Machine Vision V*, pages 104–112, November 1990.

[20] W. Kim and A. Kak. 3D Object Recogntion Using Bipartite Matching Embedded in Discrete Relaxation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 13(3):224–251, March 1991.

[21] R. Kumar and R. Hanson. Analysis of different robust methods for pose estimation. In *Proc. of IEEE Workshop on Robust Computer Vision*, October 1990.

[22] C. Lee and R. Haralick. Exterior orientation from line-to-line correspondences – a bayesian approach. In *Proc. of the IEEE Computer Vision and Pattern Recogntion*, June 1993.

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

[24] R. Mohan and R. Nevatia. Using perceptual organization to extract 3d structures. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11(11):1121 –1139, November 1989.

[25] M. Oshima and Y. Shirai. Object recognition using three-dimensional information. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 5(4):353–361, July 1983.

[26] F. Stein and G. Medioni. Structural indexing: Efficeint 3d object recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 14(2):125 – 145, February 1992.

[27] H. Wolfson. Model based object recognition by geometric hashing. In *First Europ. Conf. on Comp. Vis.*, pages 526–536, 1990.