Toward Selecting and Recognizing Natural Landmarks *

Erliang Yeh
erliang@klaupacius.eng.yale.edu
Center for Systems Science
Department of Electrical Engineering
Yale University
New Haven, CT 06520-8267

Abstract

Landmarks are often used as a basis for mobile robot navigation. In this paper, we consider the problem of automatically selecting from a set of 3D features the subset which is most likely to be recognized from noisy monocular image data and is least likely to be confused with any of the other groups of features. Assuming perspective projection, real valued recognition functions are constructed for a set of features. The value returned from such functions are invariant to changes of viewpoint and can be evaluated directly from image measurements without prior knowledge of the position and orientation of the camera. With image noise, the recognition function no longer evaluates to a constant value. Because of the possibility of false matches, a Bayes detector is used to determine the optimal range of values of the recognition function that will be accepted as image features of the model. The model with the lowest Bayes cost is selected as the most distinguishable landmark. We show implementation results for real 3D objects. Some issues, improvements and extensions to the method are discussed.

*This work was supported by the Office of Naval Research under grant N00014-93-1-0305 and the National Science Foundation under Grant NYI-IRI-9257990. An earlier version of this paper was submitted to 1995 IEEE International Conference on Intelligent Robots and Systems, coauthored by E. Yeh and D. Kriegman.

1 Introduction

There have been two approaches to mobile robot navigation in the literature: reconstructionist versus reactive. In the more traditional reconstuctionist approach, sensor information (stereo vision, motion, LIDAR, sonar, etc.) is used to construct a three dimensional model or map of the robot's environment [2, 7, 14, 17]. In this approach a great deal of effort is required to maintain a consistent (and hopefully accurate) representation of the geometry of the world [5, 8, 22, 29]. On the other hand, the reactive paradigm, initially championed by Brooks [4] and adopted by many others [1, 25], bases robot behavior more directly on immediate sensor data and less on a stored representation. In particular explicit, large scale reconstruction is avoided because as argued by proponents, the world is not static, it is difficult to maintain a consistent representation, and perhaps more importantly, it is unnecessary for most navigation tasks.

In the our own work, we have developed algorithms for systematically exploring a bounded two dimensional configuration space in search of a recognizable object [30, 31]. A prototypical task for an indoor mobile robot operating in an office setting might be fetching output from a printer. Clearly the robot must be able to recognize the printer when it is in sight. In addition to recognizing its goal, the robot takes advantage of objects that it can recognize along the way. As a byproduct, the algorithm constructs a "topological representation" of the environment akin to a level of Kuiper's spatial semantic hierarchy [19]. The representation essentially encodes which recognizable objects are visible in the vicinity of a given recognizable object, and this leads to a natural graph structure. A robot can execute a plan, defined by a path through this graph, using a combination of boundary following and "visual servoing" to approach the recognizable object. A planned path is represented much like a person's description of a route (*e.g.* go down main street until you see the traffic light and turn left, then turn right at the gas station) rather than a trajectory in some fixed, absolute coordinate system, *e.g.* $[x(t), y(t), \phi(t)]$. Exploration is then cast as the process of learning this graph and terminating when the recognizable object has been found. As a byproduct, the learned graph can be used for future navigation tasks. Note that this is not a quantitative reconstruction of the geometric structure of the environment but instead encapsulates the

qualitative relationship of recognizable objects. The graph can be augmented with metric information (e.g. distances between objects) allowing shorter routes to be planned.

The exploration/navigation algorithm described above has been implemented on our mobile robot [31], and the focus of this work as to show *how* object recognition could be used to solve navigation and exploration problems rather than using reconstruction. The actual problem of object recognition was trivialized by tacking recognizable targets (essentially bar codes) on objects; these targets are easily recognized even in cluttered scenes. One obvious approach to using natural objects rather than artificial landmarks would be to store some 3D model of a set of objects that the robot is likely to encounter and use one of the established recognition techniques such as alignment [13, 15], interpretation trees [9, 10], geometric invariance [26], aspect graphs [3, 16, 20, 27] or geometric hashing [34]. While prior models are useful for describing the destination, such an approach is going be ineffective during the course of navigation when the robot encounters many unmodelled objects. Instead, the robot should be able to learn about the new objects that it encounters and retain models of those objects that are useful for the task. Besides our own work, landmarks have been critical to many other approaches to navigation [18, 21, 23].

In this paper, we consider the problem of recognizing and learning about perceptually salient objects or landmarks from image data. Thus, a robot would not have to be preprogrammed with CAD-like models of important objects and instead would learn from what it encounters. What the robot uses a landmark will be driven by the statistical distribution of objects and features that it encounters in the world rather than some prior set of preprogrammed models.

The goal of identifying and later recognizing perceptually distinctive objects (also termed landmarks) can be cast as the following problem: *Given a set of features, select a subset of these features which in a monocular image is most likely to be recognized and least likely to be confused with any of the other group of features.* Here, we assume landmarks are selected from a set of viewpoint independent 3D features (e.g. points or lines) that are indistinguishable; that is they cannot be differentiated by local geometry, color, or texture nor can they be distinguished by adjacency information (e.g. connectivity by edges). If such information were available, it would naturally simplify the resulting combinatorics and

improve accuracy.

We take the following approach: From a set of 3D features, a subset of the features becomes a hypothetical landmark model. For this set of features, a recognition function can be constructed which evaluates to zero for any noiseless image of these features. Applying this function to actual image data, a set of features is taken to be an instance of the model when the function evaluates to zero. Because of image noise, it will not evaluate precisely to zero, and a range of values (presumable about zero) must be accepted. Knowing the probability distribution of image measurements, an optimal range can be selected based on a Bayes detector. Furthermore, the probabilities of mistaking some other object as a landmark (false positives) or missing a landmark (false negatives) can be computed. For a set of hypothetical landmarks, the one which minimizes the Bayes cost can be selected as the most salient landmarkww and used for robot navigation.

In this paper, we simplify the problem in the following way: we assume that the mobile robot only travels along a horizontal ground plane, and the only features considered are vertical lines. Together, this allows us to reduce the problem to using point features in the plane. We assume Gaussian image noise, though other models could be employed. We also assume that the 3D features are visible from any viewpoint within a certain distance (i.e., no occlusion). Taken together, these assumptions allow for tractable formulation. Future work will include methods for relaxing some of these assumptions to a richer set of features, more realistic noise models, and using representations like aspect graphs to handle occlusion.

The paper is organized as follows: In section 2, the world model and recognition functions are introduced. We then develop the probability densities for the result of applying the recognition function to noisy data, and a method for selecting the most distinguishable landmark using a Bayesian criterion is established in section 3. The approach has been implemented, and in section 4 we consider the result of applying this method. Finally, we conclude with a discussion of the method and some future directions.

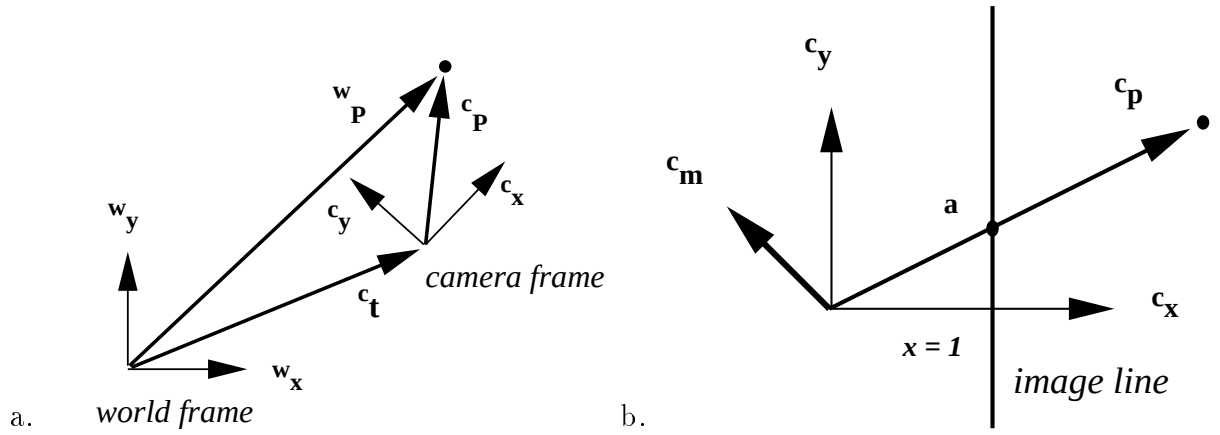


Figure 1: a. The frames and vectors associate with two views of a point. b. The perspective projection imaging model.

2 Recognition Functions and Model-based invariants

Recently, Weinshall introduced the notion of "model-based invariants" for object recognition [33]. From a set of m 3D features called the model $\mathcal{M}$, a real valued recognition function $\mathcal{I}(\mathbf{a})$ can be constructed where $\mathbf{a}$ is a vector of the image measurements. The recognition function $\mathcal{I}(\mathbf{a})$ evaluates to zero for any image of the model $\mathcal{M}$. Thus, given an image with n features, an algorithm for recognizing $\mathcal{M}$ is to choose all $\binom{n}{m}$ subsets of m features and evaluate $\mathcal{I}(\mathbf{a})$. The subset of m features which minimizes $|\mathcal{I}(\mathbf{a})|$ is considered to be the recognized object.

In this paper, we assume that the robot moves on a horizontal ground plane and that the camera is modeled by perspective projection. As in [17], note that for a camera whose optical axis is parallel to the ground plane, the image of vertical 3D lines will be vertical. Using vertical line segments as features, and projecting both the features and the image plane onto the ground plane, the problem can be modelled in two dimensions. The features project to points in the plane, the camera position is given by one orientation and two translation parameters, and the image plane can be considered an image line.

As shown in Figure 1.a, define a coordinate system attached to the camera's optical center with the x-axis in the direction of the optical axis. Given the coordinates of a point in the world frame ${}^{w}\mathbf{p}_i = ({}^{w}x_i , {}^{w}y_i)$, the coordinates of the point in the camera frame are

given by:[1]

$$
\begin{aligned}
{}^{c}\mathbf{p}_i &= {}^{c}_{w}\mathbf{R}\ {}^{w}\mathbf{p}_i + {}^{c}\mathbf{t}\,. \\
\begin{bmatrix} {}^{c}x_i \\ {}^{c}y_i \end{bmatrix} &= \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta\cos\theta \end{bmatrix} \begin{bmatrix} {}^{w}x_i \\ {}^{w}y_i \end{bmatrix} + \begin{bmatrix} {}^{c}t_x \\ {}^{c}t_y \end{bmatrix}
\end{aligned} \tag{1}
$$

where ${}^{c}_{w}\mathbf{R}$ is a 2D rotation matrix of the world frame relative to the camera frame, and ${}^{c}\mathbf{t}$ is the translation vector in camera frame.

Assuming a camera with unit focal length, the image line (projection onto the ground plane of the image plane) is located at $x = 1$ in the camera frame. Let a_i be the image measurements of ${}^{c}\mathbf{p}_i$, then

$$
a_i = \frac{{}^{c}y_i}{{}^{c}x_i} = \frac{{}^{w}y_i\ \cos\theta - {}^{w}x_i\ \sin\theta + {}^{c}t_y}{{}^{w}x_i\ \cos\theta + {}^{w}y_i\ \sin\theta + {}^{c}t_x}. \tag{2}
$$

As shown in Figure 1.b, from an image measurement a_i of a 2D point ${}^{c}\mathbf{p}_i$, we know that ${}^{c}\mathbf{p}_i$ lies on a ray defined by the optical center and the point $(a_i, 1)$. Considering the vector ${}^{c}\mathbf{m}_i = [a_i, -1]^t$ to be a vector that is orthogonal to this ray, we can derive the following constraint in the camera coordinate system:

$$
{}^{c}\mathbf{m}_i \cdot {}^{c}\mathbf{p}_i = 0 \tag{3}
$$

Expanding the above equation and expressing the coordinates of ${}^{c}\mathbf{p}_i$ in the world frame, we can construct an equation in three variables $(\theta,\ {}^{c}t_x\ ,\ {}^{c}t_y\ )$,

$$
(a_i\ {}^{w}x_i\ -\ {}^{w}y_i\ )\cos\theta + (a_i\ {}^{w}y_i\ +\ {}^{w}x_i\ )\sin\theta + a_i\ {}^{c}t_x\ -\ {}^{c}t_y\ = 0. \tag{4}
$$

Since each measurement provides one constraint on the values of $(\theta,\ {}^{c}t_x\ ,\ {}^{c}t_y\ )$, the three variables $(\theta,\ {}^{c}t_x\ ,\ {}^{c}t_y\ )$ in Equation (4) can be determined using three points and their images. For four points, we can construct a model-based recognition function $\mathcal{I}(\mathbf{a})$, where $\mathbf{a} = (a_1, a_2, a_3, a_4)$ is a vector of the image measurements in camera frame. Without loss of generality, we can let ${}^{w}\mathbf{p}_1 = (0, 0)$ and ${}^{w}\mathbf{p}_2 = (1, 0)$ by translating, rotating and scaling the

[1] To represent the coordinates of a vector, we follow the notation established by Craig [6]; the leading superscript indicates the frame in which the coordinates are expressed. Premultiplying the coordinates of a vector written in frame w by a rotation matrix ${}^{c}_{w}\mathbf{R}$ yields the coordinates in frame c.

four points. The recognition function is then of the form:

$$\begin{aligned}\mathcal{I}(\mathbf{a}) = {a_1}^2(k_1 + k_9a_2 + k_{10}a_3 + k_{11}a_4 + k_{18}a_2a_3 + k_{19}a_2a_4 + k_{20}a_3a_4) + {a_2}^2(k_2 + k_8a_1 + k_{12}a_3 \\ + k_{13}a_4 + k_{21}a_1a_3 + k_{22}a_1a_4 + k_{23}a_3a_4) + a_1(k_4a_3 + k_5a_4) + a_2(k_6a_3 + k_7a_4) \\ + a_1a_2(k_3 + k_{14}a_3 + k_{15}a_4) + a_3a_4(k_{16}a_1 + k_{17}a_2 + k_{24}a_1a_2). \qquad (5)\end{aligned}$$

where

$k_1 = k_{23} = x_4y_3 - x_3y_4 + y_3 - y_4$

$k_2 = k_{20} = x_4y_3 - x_3y_4$

$k_3 = -k_{21} = k_{24} = -2(x_4y_3 - x_3y_4) - y_3 + y_4$

$k_4 = -k_6 = -k_{19} = k_{22} = y_4$

$k_5 = -k_7 = -k_{18} = k_{21} = -y_3$

$k_8 = -k_9 = -k_{16} = k_{17} = -x_3 + x_4$

$k_{10} = -k_{13} = -y_3y_4 - x_3x_4 - x_3$

$k_{11} = -k_{12} = y_3y_4 + x_3x_4 + x_4$

$k_{14} = -k_{15} = 2(y_3y_4 + x_3x_4) + x_3 + x_4$

The coefficients k_i are constants determined by the world coordinates of the four model points. Note that $\mathcal{I}(\mathbf{a})$ is a quartic polynomial in the image measurements a_i. The value of the recognition function $\mathcal{I}(\mathbf{a})$ is not affected by changes in viewpoint. In our case, $\mathcal{I}(\mathbf{a})$ evaluates to zero for all image views if there is no noise.

By construction, the function $\mathcal{I}(\mathbf{a})$ is independent of the coordinate system used to specify p_i. Any two point sets that differ by a similarity transformation lead to the same recognition function, and so they are indistinguishable under $\mathcal{I}(\mathbf{a})$. Two sets of points that differ by a reflection will also be indistinguishable.

The goal of identifying and recognizing distinguishable landmarks from a set of 3D features is then simplified to the following 2D problem: Given a set of 2D points, generate model-based invariant recognition functions for all of the four-point models and select the one with the lowest Bayes cost as the most recognizable landmark. We now consider the selection problem.

3 Selecting distinguishable landmarks

If there were no measurement noise or detector bias, every instance of a model would be correctly recognized; the only falsely identified or missed landmarks would arise either from objects that are equivalent to the model up to some transformation or would occur from an accidental viewpoint. With image noise, the situation is different; the recognition function will no longer evaluate to precisely zero, and so a range $\mathcal{R}$ of values is employed. If $\mathcal{I}(\mathbf{a}) \in \mathcal{R}$, then $\mathbf{a}$ is considered to arise from an instance of model $\mathcal{M}$. Two similar 3D objects are likely to be indistinguishable from many viewpoints since their images will be similar; consequently for both objects, $\mathcal{I}(\mathbf{a})$ may fall within $\mathcal{R}$. To use a Bayesian approach for selecting the landmark that is most recognizable in noisy image data from the majority of viewpoints, we first need to find the probability distribution of the recognition function $p(\mathcal{I} \mid \mathcal{M}, \mathbf{v})$ over the set of viewpoints for which the features are visible.

3.1 Probability distribution for one viewpoint

First, let us consider the distribution of $\mathcal{I}(\mathbf{a})$ from a single viewpoint when $\mathbf{a}$ is corrupted by noise. For a model $\mathcal{M}$, the ideal image measurements $\mathbf{a}$ from a particular viewpoint $\mathbf{v} = (t_x, t_y, \theta)$ can be expressed as $\mathbf{a}(\mathcal{M}, \mathbf{v})$ as given in Equation (2). We assume that image measurements are corrupted by additive Gaussian noise (zero mean, a known constant variance σ), and that the noise associated with each measurement is independent. With noise, we have

$$\tilde{\mathbf{a}} = \mathbf{a}(\mathcal{M}, \mathbf{v}) + \mathbf{n}$$

w where $\mathbf{n}$ is a vector of m independent, zero mean, Gaussian random variables, each with variance σ.

The result of applying the recognition function to $\tilde{\mathbf{a}}$ is another random variable $\mathcal{I}(\tilde{\mathbf{a}})$. The probability density $p(\mathcal{I} \mid \mathcal{M}, \mathbf{v})$ could be computed using $\mathcal{I}(\mathbf{a})$, $\mathbf{a}(\mathcal{M}, \mathbf{v})$, and the known statistics of $\mathbf{n}$. However, because $\mathcal{I}(\mathbf{a})$ is nonlinear, $\mathcal{I}(\tilde{\mathbf{a}})$ will not be zero mean and will not have a normal distribution. It appears to be problematic to compute $p(\mathcal{I} \mid \mathcal{M}, \mathbf{v})$ analytically, and even if it can be found it is cumbersome. Therefore we will approximate the probability density of $(\mathcal{I} \mid \mathcal{M}, \mathbf{v})$ by a Gaussian and retain the first two moments of $(\mathcal{I} \mid \mathcal{M}, \mathbf{v})$. We now

compute these two moments.

Let m denote the power of the Gaussian noise n_i, then the moments of n_i are

$$E\left\{{n_i}^m\right\} = \begin{cases} 0 & \text{m is odd} \\ 1 \cdot 3 \ldots (m-1)\sigma^n & \text{m is even} \end{cases} \tag{6}$$

From the moments of n_i, the moments of $\tilde{a}_i$ are given by:

$$\begin{aligned}
E\left\{\tilde{a}_i\right\} &= E\left\{a_i + n_i\right\} = a_i \\
E\left\{\tilde{a}_i^2\right\} &= E\left\{{a_i}^2 + 2a_i n_i + {n_i}^2\right\} = {a_i}^2 + \sigma^2 \\
E\left\{\tilde{a}_i^3\right\} &= E\left\{{a_i}^3 + 3{a_i}^2 n_i + 3a_i {n_i}^2 + {n_i}^3\right\} = {a_i}^3 + 3a_i\sigma^2 \\
E\left\{\tilde{a}_i^4\right\} &= E\left\{{a_i}^4 + 4{a_i}^3 n_i + 6{a_i}^2 {n_i}^2 + 4{n_i}^3 a_i + {n_i}^4\right\} = {a_i}^4 + 6{a_i}^2\sigma^2 + 3\sigma^4
\end{aligned}$$

It is easy to show that if the random variables $x_1, \ldots, x_n$ are independent, the random variables $y_1 = f_1(x_1), \ldots, y_n = f_n(x_n)$ are also independent. Since the n_i's are independent, we know that $\tilde{\mathbf{a}}$ is a vector of m independent random variables. Also from the independence of n_i, if $g(\tilde{a}_i)$ is a function of $\tilde{a}_i$, then we have

$$E\left\{g(\tilde{a}_i)g(\tilde{a}_j)\right\} = E\left\{g(\tilde{a}_i)\right\} \cdot E\left\{g(\tilde{a}_j)\right\}. \tag{7}$$

Now we can compute the moments of the recognition function for a specific viewpoint when there is Gaussian image noise using the above results.

The mean of $(\mathcal{I} \,|\, \mathcal{M}, \mathbf{v})$ is

$$\begin{aligned}
\eta(\tilde{\mathbf{a}}) = E\{\mathcal{I}(\tilde{\mathbf{a}})\} = \mathcal{I}\,(\mathbf{a}) + [&k_1 + k_2 + k_8 a_1 + k_9 a_2 + (k_{10} + k_{12})a_3 + (k_{11} + k_{13})a_4 \\
&+ k_{18} a_2 a_3 + k_{19} a_2 a_4 + (k_{20} + k_{23})a_3 a_4 + k_{21} a_1 a_3 + k_{22} a_1 a_4]\sigma^2
\end{aligned} \tag{8}$$

where k_i are coefficients of $\mathcal{I}(\mathbf{a})$ given in Equation (5).

To compute the variance, we first expand $\mathcal{I}(\tilde{\mathbf{a}})$,

$$\begin{aligned}
\mathcal{I}(\tilde{\mathbf{a}}) - \eta(\tilde{\mathbf{a}}) = c_{\,1} &+ c_2 n_1 + c_3 n_2 + c_4 n_3 + c_5 n_4 + c_6 n_1 n_1 + c_7 {n_2}^2 + c_8 n_1 n_2 + c_9 n_1 n_3 \\
&+ c_{10} n_1 n_4 + c_{11} n_2 n_3 + c_{12} n_2 n_4 + c_{13} n_3 n_4 + c_{14} {n_1}^2 n_2 + c_{15} {n_1}^2 n_3 \\
&+ c_{16} {n_1}^2 n_4 + c_{17} n_1 {n_2}^2 + c_{18} {n_2}^2 n_3 + c_{19} {n_2}^2 n_4 + c_{20} n_1 n_2 n_3 + c_{21} n_1 n_2 n_4 \\
&+ c_{22} n_1 n_3 n_4 + c_{23} n_2 n_3 n_4 + k_{18} {n_1}^2 n_2 n_3 + k_{19} {n_1}^2 n_2 n_4 + k_{21} {n_1}^2 n_3 n_4 \\
&+ k_{22} n_1 {n_2}^2 n_3 + k_{23} n_1 {n_2}^2 n_4 + k_{24} {n_2}^2 n_3 n_4 + k_{20} n_1 n_2 n_3 n_4.
\end{aligned} \tag{9}$$

where

$$
\begin{aligned}
c_{14} &= k_9 + k_{18}a_3 + k_{19}a_4, c_{15} = k_{10} + k_{18}a_2 + k_{20}a_4 \\
c_{16} &= k_{11} + k_{19}a_2 + k_{20}a_3, c_{17} = k_8 + k_{21}a_3 + k_{22}a_4 \\
c_{18} &= k_{12} + k_{21}a_1 + k_{23}a_4, c_{19} = k_{13} + k_{22}a_1 + k_{23}a_3 \\
c_{20} &= k_{14} + 2(k_{18}a_1 + k_{21}a_2) + k_{24}a_4, c_{21} = k_{15} + 2(k_{19}a_1 + k_{22}a_2) + k_{24}a_3 \\
c_{22} &= k_{16} + 2k_{20}a_1 + k_{24}a_2, c_{23} = k_{17} + 2k_{23}a_2 + k_{24}a_1 \\
c_6 &= k_1 + c_{14}a_2 + k_{10}a_3 + a_4(k_{11} + k_{20}a_3) \\
c_7 &= k_2 + c_{17}a_1 + k_{12}a_3 + a_4(k_{13} + k_{23}a_3) \\
c_8 &= k_3 + 2c_{14}a_1 + 2c_{17}a_2 + k_{14}a_3 + a_4(k_{15} + k_{24}a_3) \\
c_9 &= k_4 + 2k_{10}a_1 - a_2(k_{21}a_2 + c_{20}) + k_{16}a_4 \\
c_{10} &= k_5 + 2k_{11}a_1 - a_2(k_{22}a_2 + c_{21}) + k_{16}a_3 \\
c_{11} &= k_6 + a_1(k_{14} + k_{18}a_1 + k_{24}a_4) + 2c_{18}a_2 + k_{17}a_4 \\
c_{12} &= k_7 + a_1(k_{15} + k_{19}a_1 + k_{24}a_3) + 2c_{19}a_2 + k_{17}a_3 \\
c_{13} &= -a_1(k_{20}a_1 + c_{22}) + a_2(k_{17} + k_{23}a_2), c_1 = -(c_6 + c_7)\sigma^2 \\
c_2 &= 2c_6a_1 + a_2(k_3 + k_{14}a_3 + k_{15}a_4 + c_{17}a_2) + a_3(k_4 + k_{16}a_4 + k_{24}a_2a_4) + k_5a_4 \\
c_3 &= a_1(k_3 + k_{14}a_3 + k_{15}a_4 + c_{14}a_1) + 2a_2(k_2 + k_{12}a_3 + k_{13}a_4 + c_{17}a_1) + a_3(k_6 + c_{23}a_4) + k_7a_4 \\
c_4 &= a_1(k_4 + k_{14}a_2 + k_{16}a_4 + c_{15}a_1) + a_2(k_6 + k_{17}a_4 + k_{24}a_1a_4 + c_{18}a_2) \\
c_5 &= a_1(k_5 + k_{15}a_2 + k_{16}a_3 + c_{16}a_1) + a_2(k_7 + k_{17}a_3 + k_{24}a_1a_3 + c_{19}a_2)
\end{aligned}
\tag{10}
$$

From Equation 9, we can compute the variance of $(\mathcal{I} \mid \mathcal{M}, \mathbf{v})$,

$$\sigma^2(\tilde{\mathbf{a}}) = E\left\{[\mathcal{I}(\tilde{\mathbf{a}}) - \eta(\tilde{\mathbf{a}})]^2\right\} = q_1 + q_2\sigma^2 + q_3\sigma^4 + q_4\sigma^6 + q_5\sigma^8$$

where

$$
\begin{aligned}
q_1 &= {c_1}^2, q_2 = {c_2}^2 + {c_3}^2 + {c_4}^2 + {c_5}^2 + 2c_1(c_6 + c_7) \\
q_3 &= 3({c_6}^2 + {c_7}^2) + {c_8}^2 + {c_9}^2 + c_{10}^2 + c_{11}^2 + c_{12}^2 + c_{13}^2 + 2c_2c_{17} + c_3c_{14} \\
&\quad + c_4(c_{15} + c_{18}) + c_5(c_{16} + c_{19}) + c_6c_7
\end{aligned}
$$

$$q_4 = 3({c_{14}}^2 + {c_{15}}^2 + {c_{16}}^2 + {c_{17}}^2 + {c_{18}}^2 + {c_{19}}^2) + {c_{20}}^2 + {c_{21}}^2 + {c_{22}}^2 + {c_{23}}^2$$
$$+ 2(c_9c_{26} + c_{10}c_{27} + c_{11}c_{24} + c_{12}c_{25} + c_{13}(c_{28} + c_{30}) + c_{15}c_{18} + c_{16}c_{19}$$
$$q_5 = 3(c_{24}^2 + {c_{25}}^2 + {c_{26}}^2 + {c_{27}}^2 + {c_{28}}^2 + {c_{30}}^2) + {c_{29}}^2 + 2c_{28}c_{30} \tag{11}$$

Note that q_i's are functions of $\mathbf{a}$.

3.2 Probability distribution for all viewpoints

The computation of section 3.1 provides the density of $\mathcal{I}(\tilde{\mathbf{a}})$ from only one viewpoint. Now, we can consider the distribution of $\mathcal{I}(\tilde{\mathbf{a}})$ taken over the range of viewpoints $\mathcal{V}$ from which the features are visible. Since we assumed that measurement noise $\mathbf{n}$ is independent and white, the probability density for model $\mathcal{M}$ is:

$$p(\mathcal{I} \,|\, \mathcal{M}) = \int_{\mathbf{v}\in\mathcal{V}} p(\mathcal{I} \,|\, \mathcal{M}, \mathbf{v}) p(\mathbf{v}) d\mathbf{v} \tag{12}$$

where $p(\mathbf{v})$ is the likelihood of the camera being located at viewpoint $\mathbf{v}$.

Since we assumed that measurement noise $\mathbf{n}$ is independent and white, the mean and variance of $(\mathcal{I} \,|\, \mathcal{M})$ can be computed from

$$\bar{\eta}(\tilde{\mathbf{a}}) = \int \int_{\mathbf{v}\in\mathcal{V}} \mathcal{I}\, p(\mathcal{I} \,|\, \mathcal{M}, \mathbf{v}) p(\mathbf{v}) d\mathbf{v}\, d\mathcal{I} = \int_{\mathbf{v}\in\mathcal{V}} \eta(\tilde{\mathbf{a}})\, p(\mathbf{v}) d\mathbf{v}. \tag{13}$$

$$\bar{\sigma}^2(\tilde{\mathbf{a}}) = E\{\mathcal{I}^2\} - \bar{\eta}^2 = \int \int_{\mathbf{v}\in\mathcal{V}} \mathcal{I}^2\, p(\mathcal{I} \,|\, \mathcal{M}, \mathbf{v}) p(\mathbf{v}) d\mathbf{v}\, d\mathcal{I} - \bar{\eta}^2$$
$$= \int_{\mathbf{v}\in\mathcal{V}} E\{(\mathcal{I} \,|\, \mathcal{M}, \mathbf{v})^2\} p(\mathbf{v}) d\mathbf{v} - \bar{\eta}^2 = \int_{\mathbf{v}\in\mathcal{V}} (\sigma^2 + \eta^2) p(\mathbf{v}) d\mathbf{v} - \bar{\eta}^2. \tag{14}$$

Thus, the average and variance of $p(\mathcal{I} \,|\, \mathcal{M})$ over all viewpoints is given by integrating the moments (e.g. $E\{p(\mathcal{I} \,|\, \mathcal{M})\}$) with respect to θ, t_x, and t_y for all viewpoints within the visible area $\mathcal{V}$. Supposing that the observer is equally likely to be at any viewpoint, then $p(\mathbf{v}) = \frac{1}{\int_{\mathbf{v}\in\mathcal{V}} d\mathbf{v}}$, then

$$\bar{\eta}(\tilde{\mathbf{a}}) = \frac{\int_{\mathbf{v}\in\mathcal{V}} \eta(\tilde{\mathbf{a}}) d\mathbf{v}}{\int_{\mathbf{v}\in\mathcal{V}} d\mathbf{v}}, \quad \text{and} \quad \bar{\sigma}^2(\tilde{\mathbf{a}}) = \frac{\int_{\mathbf{v}\in\mathcal{V}} (\sigma^2 + \eta^2) d\mathbf{v}}{\int_{\mathbf{v}\in\mathcal{V}} d\mathbf{v}} - \bar{\eta}^2.$$

3.3 Viewpoint space

There is a set of viewpoints for which all of the features in a model are visible. This set depends on camera resolution, the field of view of the camera, and possible occlusion by

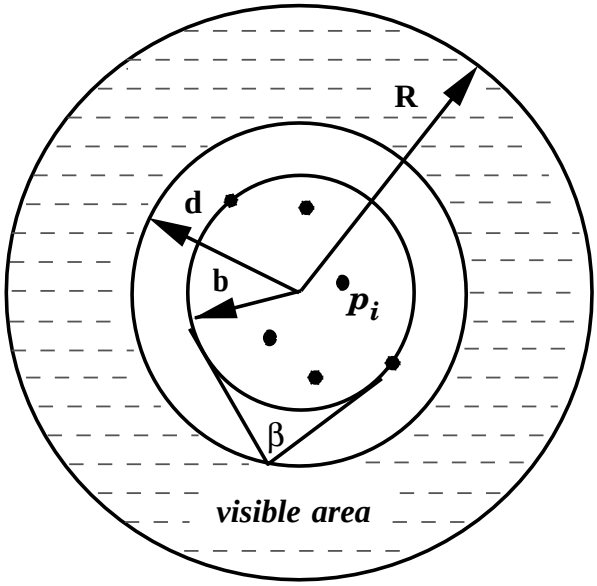


Figure 2: The visible area of a camera

other surfaces in the scene. Here, we will not be concerned with possible occlusion by opaque objects, but will handle the other two issues.

First, let β denote the field of view of the camera. If all of the feature points can be contained within a circle of radius b as shown in figure 2, then all of the features are visible for all viewpoints outside of a circle of radius $d = \frac{b}{\sin\left(\frac{\beta}{2}\right)}$. Because of the limited field of view of the camera, the model will only be visible from an interval of orientation. Expressing the location of the camera center in polar coordinates (r, α) as $t_x = r\cos\alpha$ and $t_y = r\sin\alpha$, the range of camera orientations for which all of the features are visible is given by:

$$\theta \in (\theta_{min}(r), \theta_{max}(r)) = (\alpha + \pi - \frac{\beta}{2} + \phi, \alpha + \pi + \frac{\beta}{2} - \phi)$$

where $\phi = \arctan\left(\frac{d}{r}\right)$ is the angular deviation of the camera orientation at a distance r from the center of the circle. Thus, for any camera center outside of the circle of radius d and orientation within $(\theta_{min}, \theta_{max})$, all of the feature points will be within the field of view; therefore any model $\mathcal{M}_i$ will also be visible. To account for the finite resolution of the camera, we assume that all features contained within a circle of radius b will be visible from any viewpoint inside a circle of radius R. Thus, the *visible area* is taken to be an annulus.

From the change of coordinates, we have $d\theta \cdot dt_x \cdot dt_y = r \cdot d\theta \cdot dr \cdot d\alpha$. The spatial mean

of $p(\mathcal{I} \mid \mathcal{M})$ is

$$\bar{\eta}(\tilde{\mathbf{a}}) = \frac{\int_0^{2\pi} \int_d^R \int_{\theta_{min}(r)}^{\theta_{max}(r)} \eta(\tilde{\mathbf{a}})\, r d\theta dr d\alpha}{\int_0^{2\pi} \int_d^{R(\alpha)} \int_{\theta_{min}}^{\theta_{max}} r d\theta dr d\alpha}$$

The spatial variance of $p(\mathcal{I} \mid \mathcal{M})$ is

$$\bar{\sigma}^2(\tilde{\mathbf{a}}) = \frac{\int_0^{2\pi} \int_d^R \int_{\theta_{min}(r)}^{\theta_{max}(r)} (\sigma^2 + \eta^2)\, r d\theta dr d\alpha}{\int_0^{2\pi} \int_d^{R(\alpha)} \int_{\theta_{min}}^{\theta_{max}} r d\theta dr d\alpha} - \bar{\eta}^2$$

Unfortunately, the mean $\bar{\eta}(\tilde{\mathbf{a}})$ and the variance $\bar{\sigma}^2(\tilde{\mathbf{a}})$ cannot be integrated analytically, and so they can be approximated by computing the finite sum with suitably fine sampling. Given the spatial probability distribution of $p(\mathcal{I} \mid \mathcal{M})$, we now compute the recognition interval of the model $\mathcal{M}$ using the Bayes criterion.

3.4 Computing recognition intervals

Consider what happens when $\mathcal{I}(\mathbf{a})$ is applied to noisy images of some other set of feature points $\mathcal{G}$. From $\mathcal{I}(\mathbf{a})$, $\mathcal{G}$, and the known statistics of $\mathbf{n}$, the density function $p(\mathcal{I} \mid \mathcal{G})$ can be determined for observing $\mathcal{G}$ from all viewpoints. Thus we have the distribution of applying $\mathcal{I}(\mathbf{a})$ to the correct model $\mathcal{M}$ and to an incorrect set of points $\mathcal{G}$. Typically the distribution of $p(\mathcal{I} \mid \mathcal{M})$ is nearly zero mean and has a fairly small variance, whereas $p(\mathcal{I} \mid \mathcal{G})$ is likely to have a mean that is far from zero and a rather broad distribution. That is, for a mismatch, the value of $\mathcal{I}(\mathbf{a})$ is likely to be far from zero and to vary quite a bit with viewpoint.

3.4.1 Bayes criterion

The recognition problem is to decide, based on the value returned by the recognition function from a single observation, whether or not a set of image features is identified as $\mathcal{M}$. We call hypothesis H_0 the event that image measurements are the image of $\mathcal{M}$ and the alternative hypothesis H_1 that the features do not arise from $\mathcal{M}$. There is a probabilistic description corresponding to each hypothesis. We know that either H_0 or H_1 is true. A criterion for making the decision must be selected. That is, given a recognition value $\mathcal{I}(\mathbf{a})$, which hypothesis is most probably true? The Bayes criterion can be used to determine the optimal range $\mathcal{R}$ of values of $\mathcal{I}(\mathbf{a})$ which will be accepted as images of $\mathcal{M}$.

There are two kinds of errors that can be made. One is to choose H_0 given H_1 is true (false negative), the other one is to select H_1 when H_0 is true (false positive). Depending upon the application, the consequences of each type of error may not be equally important, and so costs are assigned to each type of error. Let C_{ij} denote the cost associated with choosing hypothesis H_i when in fact hypothesis H_j is true. Without loss of generality, let $C_{00} = C_{11} = 0$ and $C_{10} > C_{00}$ and $C_{01} > C_{11}$. The Bayes criterion is to select a $\mathcal{R}$ so that the average cost will be minimized. Thus the region $\mathcal{R}$ where H_1 is chosen is [32]:

$$\mathcal{R} = \{\mathcal{I} \in \mathbb{R} : p(\mathcal{I} \mid H_1) > p(\mathcal{I} \mid H_0)\,(\frac{p(H_0)}{p(H_1)})(\frac{C_{10}}{C_{01}})\}$$

We denote H_1 as the hypothesis that $\mathcal{M}$ is present and H_0 as the hypothesis that $\mathcal{M}$ is not present. Let $\mathcal{G}_j$ denote some model other than $\mathcal{M}$. If the only features in the scene arise from the hypothetical models $\mathcal{M}$ and $\mathcal{G}_j$, then

$$\begin{aligned} p(H_1) &= p(\mathcal{M}) \\ p(H_0) &= \textstyle\sum_{j=1}^{n-1} p(\mathcal{G}_j) = 1 - p(\mathcal{M}). \end{aligned}$$

Assuming that all features and consequently all models are equally probable, then we have $p(\mathcal{M}) = \frac{1}{n}$, $p(H_0) = \frac{n-1}{n}$, and $p(\mathcal{G}_j) = \frac{1}{n-1}$. Furthermore, the conditional probabilities for the two hypotheses are given by:

$$\begin{aligned} p(\mathcal{I} \mid H_1) &= p(\mathcal{I} \mid \mathcal{M}) \\ p(\mathcal{I} \mid H_0) &= \textstyle\sum_{j=1}^{n-1} p(\mathcal{I} \mid \mathcal{G}_j) p(\mathcal{G}_j). \end{aligned}$$

Bayes' rule and the conditional densities $p(\mathcal{I} \mid \mathcal{M})$ given in Equation (12) can be used to compute $\mathcal{R}$:

$$\mathcal{R} = \{\mathcal{I} \in \mathbb{R} : p(\mathcal{I} \mid \mathcal{M}) > (\frac{\mathcal{C}_{10}}{\mathcal{C}_{01}}) \sum_{j=1}^{n-1} p(\mathcal{I} \mid \mathcal{G}_j)\}$$

The range $\mathcal{R}$ will minimize the average Bayes cost.

3.4.2 Recognition Interval

The optimal range $\mathcal{R}$ may be composed of a set of disjoint intervals. Rather than employing all of the intervals of $\mathcal{R}$, we use the single interval about the mean of $p(\mathcal{I} \mid \mathcal{M})$. In particular we denote the interval $(x_l, x_r) \subset \mathcal{R}$ such that $\bar{\eta} \in (x_l, x_r)$ as the recognition interval of model $\mathcal{M}$. Since the conditional probability density functions are differentiable, we can use

Newton's method [28] to solve the following equation to find the limits of the recognition interval (x_l, x_r).

$$f(\mathcal{I}) = p(\mathcal{I} \mid \mathcal{M}) - (\frac{\mathcal{C}_{10}}{\mathcal{C}_{01}}) \sum_{j=1}^{n-1} p(\mathcal{I} \mid \mathcal{G}_j) = 0$$

To decide the lower bound x_l, we repeat Newton's method by making several initial guesses that are smaller than the spatial mean $\bar{\eta}$ within a reasonable range to obtain a set of roots. Since the recognition interval will minimize the Bayes cost, the derivative of $f(\mathcal{I})$ should be positive at both x_l and x_r. Since there may be more than one root, we choose the one closest to the spatial mean $\bar{\eta}$ as x_l. Similarly, we repeat Newton's method by giving a few initial guesses that are larger than $\bar{\eta}$ and select the root with positive derivative and closest to $\bar{\eta}$ as the upper bound x_r.

3.5 Selecting the landmark

We are now ready to select the most salient or easily recognized constellation of features as a model from a given set of features. The n features in the set can be grouped into $l = \binom{n}{4}$ hypothetical models $\mathcal{M}_i$ with $i \in [1, \ldots, l]$ containing $m = 4$ points, and the corresponding recognition function $\mathcal{I}_i$ can be constructed. For a model $\mathcal{M}_i$, all other models $\mathcal{M}_j, i \neq j$ can be treated as $\mathcal{G}_j$, and the recognition interval $\mathcal{R}_i$ of the model $\mathcal{M}_i$ can be computed . We can then compute the total Bayes cost $\mathbf{C}_B$ using the error function. Given l models, there are $(l-1)$ mismatch models $\mathcal{G}_j$ for each model $\mathcal{M}_i$. Since we assume that $p(\mathcal{I} \mid \mathcal{M}_i)$ is a Gaussian distribution with mean $\bar{\eta}_i$ and variance $\bar{\sigma}_i^2$, the cost of false negative recognizing the model $\mathcal{M}_i$ using the recognition interval (x_l, x_r) is:

$$\mathbf{F}_n = 1 - \int_{\mathcal{I} \in \mathcal{R}_i} p(\mathcal{I} \mid \mathcal{M}_i) d\mathcal{I} = 1 - \mathbf{erf}\left(\frac{x_r - \bar{\eta}_i}{\bar{\sigma}_i}\right) + \mathbf{erf}\left(\frac{x_l - \bar{\eta}_i}{\bar{\sigma}_i}\right)$$

The cost of false positives (misidentifying something else as the model) is

$$\mathbf{F}_p = (\frac{\mathcal{C}_{10}}{\mathcal{C}_{01}}) \sum_{j=1}^{n-1} \int_{\mathcal{I} \in \mathcal{R}_i} p(\mathcal{I} \mid \mathcal{G}_j) d\mathcal{I} = (\frac{\mathcal{C}_{10}}{\mathcal{C}_{01}}) \sum_{j=1}^{n-1} \left[\mathbf{erf}\left(\frac{x_r - \bar{\eta}_j}{\bar{\sigma}_j}\right) - \mathbf{erf}\left(\frac{x_l - \bar{\eta}_j}{\bar{\sigma}_j}\right)\right]$$

where $\mathbf{erf}x$ is the error function

$$\mathbf{erf}x = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \; - \; \frac{1}{2}.$$

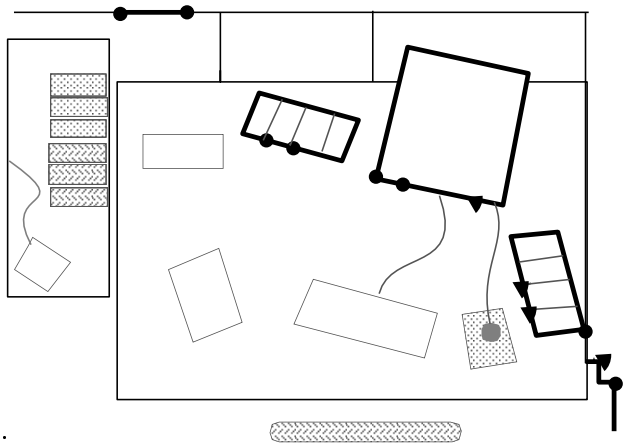
a.  b. 

Figure 3: a. An image of an office scene. b. A drawing of the scene; the landmark is selected from the set of vertical lines features drawn as darkened points, and the features selected for the landmark are shown as triangles. The recognition interval is $(-0.017228, -0.006332)$

The Bayes cost for using $\mathcal{I}$ to recognize $\mathcal{M}$ with interval (x_l, x_r) can be computed from

$$\mathbf{C}_B = \mathbf{F}_p + \mathbf{F}_n$$

The total Bayes cost $\mathbf{C}_B$ can be computed for each candidate model, and the models can be sorted according to these rates. Those models with lower average cost are more likely to be recognized from noisy image data and not confused from the majority of viewpoints. We thus select the model with the lowest Bayes cost as the most distinguishable landmark.

4 Implementation and examples

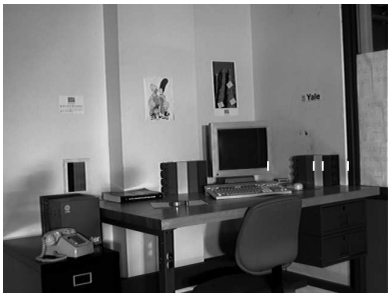
The presented approach to landmark selection has been prototyped in Common Lisp. Figure 3.a shows an image of an office, and figure 3.b shows an overhead view. A subset of 12 vertical lines in the scene were considered features, and these are indicated in figure 3.b. The landmark selection process was applied assuming that image measurements are corrupted by Gaussian noise with a standard deviation of one pixel. The optimal landmark with the lowest Bayes cost was selected according to procedure in section 3, and the features comprising this landmark are indicated by darkened triangles in figure 3.b.

Note that there are two kinds of errors which can be made, the consequence of false negative will be more important than the one of false positive since we don't want to mismatch

a. invariant=-0.016947
correctly recognized.

b. invariant=-0.017228
correctly recognized.

c. invariant=-0.015875
correctly recognized.

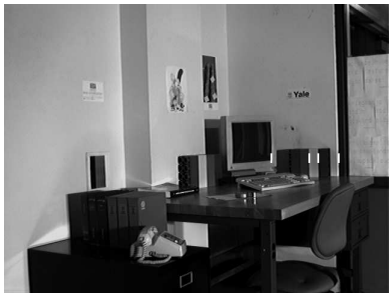


d. invariant=-0.012982
correctly recognized.

e. invariant=-0.010899
correctly recognized.

f. invariant=-0.009315
correctly recognized.

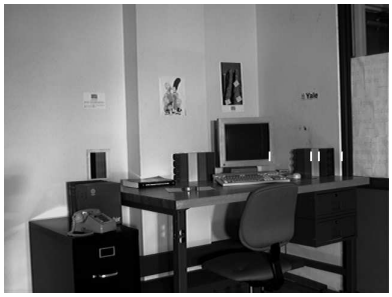
g. invariant=-0.008608
correctly recognized.



h. invariant=-0.008127
correctly recognized.

i. invariant=-0.012406
correctly recognized.

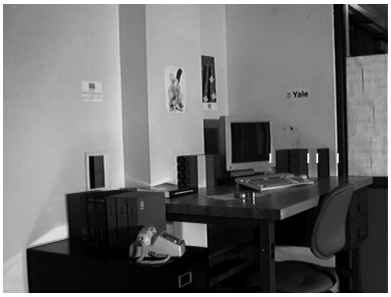
j. invariant=-0.013397
correctly recognized.



k. invariant=-0.010230
correctly recognized.

l. invariant=-0.009026
correctly recognized.

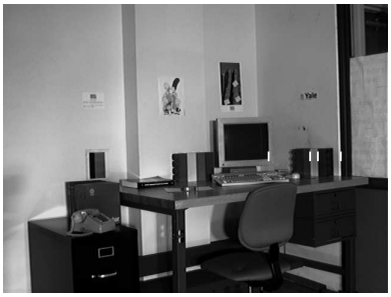
m. invariant=-0.008103
correctly recognized.



n. invariant=-0.008127
correctly recognized.

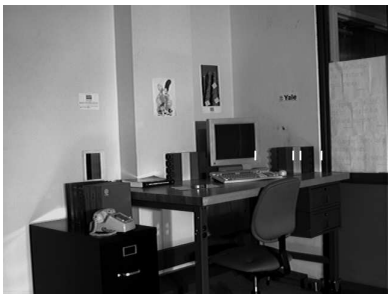
o. invariant=-0.008510
correctly recognized.

p. invariant=-0.012406
correctly recognized.

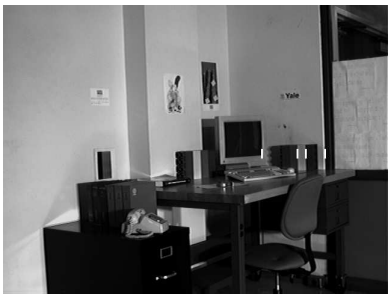


q. invariant=-0.009916
correctly recognized.

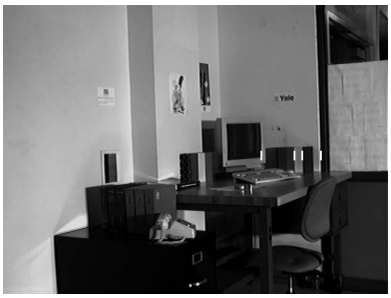
r. invariant=-0.009483
correctly recognized.



s. invariant=-0.008330
correctly recognized.



t. invariant=-0.006994
correctly recognized.



u. invariant=-0.006333
correctly recognized.

Figure 4: The images that selected landmark being found correctly and uniquely by applying the recognition function to the k automatically detected edges (between 13 and 21) and consequently $\binom{k}{4} = (715 \sim 5985)$ groups of features with ordering and gradient constraints.

t. detected 17 edges
found 2 matches

t_1. invariant=-0.012793
correct match.

t_2. invariant=-0.013129
false match.

u. detected 17 edges
found 2 matches

u_1. invariant=-0.008127
correct match.

u_2. invariant=-0.011139
false match.

Figure 5: The experimental results of finding two matches with ordering and gradient constraints by applying the recognition function to $\binom{17}{4} = 2380$ groups of features. Each image has one correct match and one false match.

v. detected 17 edges
found 2 mismatches

v_1. invariant=-0.011153
false match.

v_2. invariant=-0.014390
false match.

Figure 6: The result that selected landmark was not correctly recognized. There were two false positive matches in the image.

the correct landmark. In our implementation, the cost factor $\frac{\mathcal{C}_{01}}{\mathcal{C}_{10}}$ was assigned to be close to the number of all hypothetical landmarks. The mean value of $\mathcal{I}(\mathbf{a})$ for the selected landmark was -0.00825 and the recognition interval was $(-0.01722 - 0.00632)$. The computed average cost for this landmark was 0.04951. Note that the worst hypothesized landmark had a average cost of 0.40276.

We then attempted to recognize the selected landmark in images. A camera was moved to 24 positions covering a quarter circle at three depths, and images were digitized with a resolution of 640 by 480 pixels. The selection of an optimal landmark assumed that all features are visible and no other vertical lines were considered as additional features in the image. This is done by extracting the 12 selected vertical edges manually. Given 12 features, $frac12!(12-4)! = 11880$ groups of hypothetical landmarks can be formed. Without any constraints, the recognition function was applied to 11880 groups of features, and those falling within the recognition interval were taken as instances of landmark. There were totally 18 false positive matches and 1 false negative match found in the 24 images. The resultant Bayes cost was 0.049 which is close to the threoretical one.

Since the order of the selected vertical lines doesn't change in the image, we can use the ordering constraint to reduce matching combinatories. Given 12 features, there are $\binom{12}{4} = 495$ groups of hypothetical landmarks can be formed with ordering constraint. The recognition function was applied to 495 groups of features and the resultant Bayes cost was 0.048. The landmark was recognized in 23 out of 24 images, and it was recognized as the only landmark in 18 of those 23 images. In the other 5 images, there were 1 to 4 false matches found in addition to the correct one and resulted in tatally 11 false positive matches. The landmark was not recognized in only one image and there were two false matches found in that image.

Since the signs of the gradients of the gray levels of the four edges in the landmark and the the ones in the subsets of features must be consistant, we can apply this gradient constraint to the recognition process to improve the performance. With ordering and gradient constraints, the Bayes cost was 0.047 and the number of false positive matches was reduced from 11 to 0.

To detect the edges automatically, we apply a one dimensional edge detector across one

row of the image to extract all of the k vertical edges (between 13 and 21) crossing that line. Given k features, there are $\frac{k!}{(k-4)!} = (17160 \sim 143640)$ groups of hypothetical landmarks can be formed. With no constraints, the recognition function was applied to $17160 \sim 143640$ groups of features and the resultant Bayes cost was 0.048. There were totally 72 false positive matches and 1 false negative match were found in the 24 images. With ordering constraint, we apply the recognition function to $(715 \sim 5985)$ groups of features and the Bayes cost was 0.048. The selected landmark was correctly recognized in 23 images and 57 false positive matches were found in the 24 images. With ordering and gradient constraints, the Bayes cost was 0.047 and the number of false positive matches was reduced to 3. This results are shown in Figure 4, Figure 5 and Figure 6.

Figure 4 shows a series of 21 images for which the selected landmark was correctly and uniquely recognized by applying the recognition function to $13 \sim 21$ automatically detected vertical edges and consequently $715 \sim 5985$ groups of features with ordering and gradient constraints. The four highlighted vertical edges in the image indicate the selected landmark. Figure 5 shows the 2 images that two matches were found, each with one correct match and one false match. Figure 6 shows the image that the selected landmark was not found. There were two false positive matches in the image.

The implementation results are shown in the following atbles,

	theoretically			*manually*			*automatically*		
constraints	$\mathbf{C}_B$	F_p	F_n	$\mathbf{C}_B$	N_{fp}	N_{fn}	$\mathbf{C}_B$	N_{fp}	N_{fn}
none	0.049	0.008	0.041	0.048	282	1	0.047	1346	1
ordering	0.045	0.005	0.040	0.048	11	1	0.049	57	1
ordering and gradient				0.047	0	1	0.047	3	1

Table 1: Implementation results of selected landmark.

	theoretically			*manually*			*automatically*		
constraints	$\mathbf{C}_B$	F_p	F_n	$\mathbf{C}_B$	N_{fp}	N_{fn}	$\mathbf{C}_B$	N_{fp}	N_{fn}
none	0.402	0.045	0.347	0.388	1832	8	0.387	10243	8
ordering	0.384	0.031	0.353	0.391	116	8	0.387	437	8
ordering and gradient				0.384	41	8	0.381	102	8

Table 2: Implementation results of the bad landmark.

In both tables, $\mathbf{C}_B$ denotes the Bayes cost, F_p is the false positive cost, F_n is the false negative cost, N_{fp} denotes the number of images that the landmark is false positive recognized, and N_{fn} is the number of images that the landmark is false negative recognized. As we can see from the above tables, the low Bayes cost of the selected landmark resulted in a much larger number of correct matches than the bad landmark in all cases, that is, the selected landmark is much more distinguishable than the bad landmark.

5 Discussion

The method described is a starting point for a Bayesian approach to landmark selection. In the process, a number of assumptions and simplifications were made. Further empirical investigation is needed to determine the validity of this model for landmark selection. There are a number of issues, improvements and extensions to this basic scheme.

- First, we assumed that measurements are corrupted by additive Gaussian noise; other more realistic noise models or distributions which are more computationally attractive should be considered.

- When hypothesizing possible landmarks, all $\frac{n!}{(n-4)!}$ hypothetical groups of features were considered. This is an explosive number, and so principled means of reducing the number of hypothetical landmarks must be developed.

- We assumed that all features were visible from all viewpoints in the viewpoint space $\mathcal{V}$ when computing of $\mathcal{R}_i$. Instead, an aspect graph or similar representation could be used to determine the set of viewpoints for which the features in $\mathcal{M}_i$ are not occluded, and this could be used to compute $p(\mathcal{M}_i)$, the probability that $\mathcal{M}_i$ is visible within $\mathcal{V}$.

- The detectibility of the individual features could also be considered, perhaps as a function of viewpoint, and this could be folded into the above scheme.

- A relevant variation is to also consider selecting the most salient landmark when the coordinates of the 3D features used to define the models are noisy. This would arise when landmark selection is performed on-line.

We have presented the methodology in terms of landmark selection, but the same techniques can be applied to other object recognition problems. For example, in interpretation tree [10, 11] or alignment methods, the above analysis can be used for automatically determining the thresholds for accepting a hypothetical model, to determine termination conditions, and to order the search process through the interpretation tree. A set-based approach as opposed to a probabilistic approach to this problem was presented in [12]. An outgrowth of the above work may indicate what feature geometries and recognition functions are least sensitive to noise independent of the other features; this could lead to a method of object selection that does not require strict pairwise comparison. It may also indicate how to design artificial landmarks.

The two methods outlined above rely on having three dimensional data available for constructing the models. A very interesting extension is to consider the problem of landmark selection using only data from a single image of a scene. Between two images of an object, the epipolar constraint has to be satisfied for all corresponding features [24]. An object can be modelled by the image coordinates of a set of features from one viewpoint. A recognition function is then constructed which measures the degree to which another set of image features measured in a second image violates the epipolar constraint. Without noise, a correct correspondence will evaluate to zero, and mismatches will evaluate to some other number. This recognition function could be directly used in the above Bayes classification scheme. In practice, the epipolar constraint is likely to be too weak to be used alone. Other constraints will have to be brought to bear to determine which model is most distinguishable: These might include feature type, connectivity or adjacency relationships between pairs of features, and photometric information; this would couple some of the notions of qualitative image structure given by an aspect graph and would lead to a multiple-view representation of individual objects.

Acknowledgements

Many thanks to C.J. Taylor whose work on landmark-based mobile robot navigation started us on this path.

References

[1] R. Arkin and R. Murphy. Autonomous navigation in a manufacturing environment. *IEEE Trans. on Robotics and Automation*, 6(5):445–454, August 1990.

[2] N. Ayache and O. Faugeras. Maintaining representations of the environment of a mobile robot. *IEEE Trans. on Robotics and Automation*, 5(6):804–819, December 1989.

[3] K. Bowyer and C. R. Dyer. Aspect graphs: An introduction and survey of recent results. *Int. J. of Imaging Systems and Technology*, 2:315–328, 1991.

[4] R. A. Brooks. A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, 2(1):14–23, Mar. 1986.

[5] R. Chatila and J. Laumond. Position referencing and consistent world modeling for mobile robots. In *IEEE Int. Conf. on Robotics and Automation*, 1985.

[6] J. Craig. *Introduction to Robotics: Mechanics and Control.* Addison-Wesley, New York, 1989.

[7] J. Crowley. Navigation for an intelligent mobile robot. *IEEE Journal of Robotics and Automation*, pages 31–41, Mar. 1985.

[8] A. Elfes. Sonar-based real-world mapping and navigation. *IEEE Journal of Robotics and Automation*, 3(3):249–265, June 1987.

[9] O. Faugeras and M. Hebert. The representation, recognition, and locating of 3-D objects. *Int. J. Robot. Res.*, 5(3):27–52, Fall 1986.

[10] W. Grimson and T. Lozano-Perez. Localizing overlapping parts by searching the interpretation tree. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 9(4):469–482, 1987.

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

[12] W. E. L. Grimson and D. P. Huttenlocher. On the verification of hypothesized matches in model-based recognition. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 13(12):1201–1213, 1991.

[13] D. Huttenlocher and S. Ullman. Object recognition using alignment. In *Int. Conf. on Computer Vision*, pages 102–111, London, U.K., June 1987.

[14] A. Kosaka and A. Kak. Fast vision-guided mobile robot navigation using model-based reasoning and prediction of uncertainties. *CVGIP: Image Understanding*, 56(3):271–329, 1993.

[15] D. Kriegman and J. Ponce. On recognizing and positioning curved 3D objects from image contours. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 12(12):1127–1137, 1990.

[16] D. J. Kriegman and J. Ponce. Computing exact aspect graphs of curved objects: Solids of revolution. *Int. J. Computer Vision*, 5(2):119–135, 1990.

[17] D. J. Kriegman, E. Triendl, and T. O. Binford. Stereo vision and navigation in buildings for mobile robots. *IEEE Trans. on Robotics and Automation*, 5(6):792–803, Dec. 1989.

[18] K.-D. Kuhnert. Fusing dynamic vision and landmark navigation for autonomous driving. In *IEEE Int. Workshop on Intelligent Robots and Systems*, pages 113–119, July 1990.

[19] B. Kuipers and Y. Byun. A robot exploration and mapping strategy based on a semantic hierarchy of spatial representations. *Robotics and Autonomous Systems*, 8:47–63, 1981.

[20] A. Laurentini. The power of the aspect graph for topological discrimination of polyhedral objects. Dipartimento di automatica ed informatica, Politecnico di Torino, Corso Duca Degli Aburzzi 24, 10129 Torino Italy, Sept. 1990.

[21] A. Lazanas and J.-C. Latombe. Landmark-based robot navigation. In *Proc. Am. Assoc. Art. Intell.*, 1992.

[22] J. Leonard, H. Durrant-Whyte, and I. Cox. Dynamic map building for an autonomous mobile robot. In *IEEE Int. Workshop on Intelligent Robots and Systems*, pages 89–96, 1990.

[23] T. Levitt, D. Lawton, D. Chelberg, and P. Nelson. Qualitative navigation. In *Proc. Image Understanding Workshop*, pages 447–465, 1987.

[24] H. Longuet-Higgins. A method of obtaining the relative positions of four points from three perspective projections. *Image and Vision Computing*, 10(5):266–270, 1992.

[25] M. Mataric. Integration of representation into goal directed behavior. *IEEE Trans. on Robotics and Automation*, 8(3):304–312, June 1992.

[26] J. Mundy and A. Zisserman. *Geometric Invariance in Computer Vision.* MIT Press, Cambridge, Mass., 1992.

[27] S. Petitjean, J. Ponce, and D. Kriegman. Computing exact aspect graphs of curved objects: Algebraic surfaces. *Int. J. Computer Vision*, 9(3):231–255, 1992.

[28] W. Press, B. Flannery, S. Teukolsky, and W. Vetterling. *Numerical Recipes in C.* Cambridge University Press, 1988.

[29] R. Smith and P. Cheeseman. On the representation and estimation of spatial uncertainty. *Int. J. Robot. Res.*, 5(4):56–68, 1986.

[30] C. Taylor and D. Kriegman. Exploration strategies for mobile robots. In *IEEE Int. Conf. on Robotics and Automation*, volume 2, pages 248–253, May 1993.

[31] C. Taylor and D. Kriegman. Algorithms for vision-based exploration. In *Workshop on the Algorithmic Foundations of Robotics*, Jan. 1994.

[32] H. L. V. Trees. *Detection, Estimation, and Modulation Theory.* John Wiley and Sons, 1968.

[33] D. Weinshall. Model-based invariants in 3-D vision. *Int. J. Computer Vision*, 10(1):27–42, 1993.

[34] H. Wolfson. Model-based object recognition by geometric hashing. In *European Conf. on Computer Vision*, pages 526–536, 1990.