Machine Vision and Applications (1997) 9: 304–320

Machine Vision and Applications

A methodology for evaluation of task performance in robotic systems: a case study in vision-based localization

Péter L. Venetianer, Edward W. Large, Ruzena Bajcsy

GRASP Lab, Computer and Information Science Department, University of Pennsylvania, 3401 Walnut 301, Philadelphia, PA 19104-6228, USA
venetian@grip.cis.upenn.edu, large@grip.cis.upenn.edu, bajcsy@central.cis.upenn.edu

Abstract. We investigated the performance of an agent that uses visual information in a partially unknown and changing environment in a principled way. We propose a methodology to study and evaluate the performance of autonomous agents. We first analyze the system theoretically to determine the most important system parameters and to predict error bounds and biases. We then conduct an empirical analysis to update and refine the model. The ultimate goal is to develop self-diagnostic procedures. We show that although simple models can successfully predict some major effects, empirically observed performance deviates from theoretical predictions in interesting ways.

1 Introduction

In this paper we investigate the performance of an agent which acts using visual information in a partially unknown and changing environment in a principled way. Unlike in the empirical sciences, where experiments are performed to identify mechanisms of task performance, in the engineering sciences artificial agents are designed according to well known principles. Thus we know the contents of our "black boxes" *a priori* and can predict in detail how the agent will perform in a constrained environment. What is often less clear, however, is how the artificial agent will interact with the (very complex) real world. The environment in which the agent perceives and acts provides uncontrolled sources of variability determining the performance of the agent. Our goal, predicated by good engineering science, is to determine by empirical experiment aspects of agent-environment interaction that affect performance in systematic ways. In so doing, we hope to show the relationship between a deterministic model of the agent, systematic determinants of performance that arise from the agent-environment interaction, and those determinants of performance that are best modeled stochastically.

In order to focus on these issues we selected the task of landmark based localization for a mobile agent, the determination of position and orientation relative to a landmark (also known as pose estimation). Within the larger system, localization serves two purposes. First, we assume that the environment contains definite, distinguishable landmarks that the agent can detect. By estimating its location with respect to landmarks the agent can determine its global position using an internal map that indicates the position of each landmark. These landmarks are assumed to be stationary. Second, by identifying landmarks mounted on other mobile agents, each agent can identify others and localize itself with respect to others. In both cases the agent must determine its distance from the landmark and its orientation with respect to the normal of the landmark. This operation does not assume any representation of the environment, nor is it based on potentially unreliable odometry readings. However, this strategy can be combined with odometry readings which can regularly be recalibrated using the more precise localization strategy. The localization algorithm used here is not unique or novel, rather it serves as a means of demonstrating our experimental methodology.

Our analysis proceeds in two stages. First, we analyze the equations that the agent uses to perform the localization task. This analysis helps us to identify the most important variables in determining the performance of the agent as it interacts with the environment [1]. In addition, the analysis makes certain predictions about the performance of the localization algorithm. Next, informed by this analysis, we design an empirical experiment with which to test the performance of the agent in a realistic environment. We use an analysis of variance to evaluate our results.

By these means we address three issues. First, we attempt to verify that the agent actually performs within predicted error bounds as it interacts with the environment. Second, we determine systematic deviations from predicted performance as we manipulate various factors that are determined by the agent-environment interaction. Finally, we address the basic methodological question of how well we can model the agent-environment interaction based on physics and geometry, what factors are best explored by empirical experimentation, and what must or can be understood or modeled as stochastic processes.

The remainder of this paper is organized as follows: Sect. 2 presents the localization algorithm, Sect. 3 addresses

performance issues through a detailed sensitivity analysis of the localization algorithm, and Sect. 4 describes an experiment designed to test the performance of the system under realistic conditions and uses an analysis of variance to evaluate the results. We conclude by discussing the implications of our results for performance testing and for building more robust autonomous agents.

2 The localization algorithm

The recognition of landmarks and estimation of relative position are accomplished via a landmark based localization algorithm. The goal of the localization algorithm is to find and track stationary or moving landmarks in the environment and determine their position relative to the viewing camera. This functionality serves two purposes:

1. Determine agent's absolute location: Landmarks are placed at various static locations in the environment, identifying positions in the agent's internal map. The algorithm determines agent position relative to landmarks, thus absolute position in the environment can be determined.
2. Identify and localize other agents: Landmarks are mounted on four sides of each agent, uniquely coding the identity and side of the agent. Therefore another agent can be recognized from any direction and its position relative to the viewing camera can be determined.

This task is a variant of the well-known pose-estimation problem [2, 3, 4], using artificial landmarks with known properties [5, 6]. The task is composed of two main steps:

1. Landmark recognition: Identify the landmark and determine its size.
2. Pose estimation: Calculate the position and orientation of the agent with respect to the landmark based on agent parameters, environmental constraints, and data retrieved in the landmark recognition step.

The next two subsections describe these steps in detail.

2.1 Landmark recognition

We choose landmarks that are easily detectable, and distinguishable both from the background and from one another so as to provide robust cues for the pose estimation. We use black-and-white 2D landmarks pasted horizontally on a planar surface. The landmarks are composed of two parts, a set of vertical bars and a binary code (see Fig. 1).

The left portion of the landmark contains four equally sized vertical black bars common to all landmarks. It is used to detect the location of the landmark. It is easy to recognize and our experience shows that other objects rarely result in a similar pattern causing false alarms. The right half of the landmark represents a binary code that is unique to each landmark, therefore it can be used unambiguously to identify other agents or places in the environment. The legal binary codes were selected in such a way as to minimalize the probability of misdetection. This is a simplified version of real-life scenarios in which street signs, road signs, and so forth may serve as landmarks. However, we believe that our algorithm for vision based localization is general modulo the following assumptions [7]: (a) the markers are posted on a plane, and (b) the agent has the template for the marker, including its overall size and expected height from the ground.

The landmark recognition algorithm contains three main steps: gray-scale to black-and-white conversion, vertical bar detection, and binary code extraction. Gray-scale to black-and-white conversion is performed either using constant or adaptive thresholding. Both assume the one byte representation of a pixel. The former uses a constant threshold of 128 whereas the latter uses the average pixel value of the entire image as threshold. Absolute thresholding is faster but less robust. Vertical bar detection and binary code extraction make use of the fact that the landmarks are mounted horizontally and scan the black-and-white image horizontally line by line. Because these steps are not crucial to the experiment to be described, they are not explained in further detail here.

2.2 Pose estimation

Once a landmark is recognized on the camera image, the next step is to determine the position of the camera relative to the landmark. Relative position can then be used to determine the location of the agent within the environment, given a stationary landmark whose actual location is known. Alternatively, it can be used to determine the position of another agent relative to the camera. The landmark detection algorithm determines the relative position of the landmark in terms of the following variables, as shown in Fig. 2:

- R: the horizontal component of the distance between the center of the agent and the center of the landmark
- θ: the angle between the heading of the agent and the center of the landmark
- Γ: the angle between the heading of the agent and the plane containing the landmark

Thus, R and θ are the polar coordinates of the landmark in an agent-centered coordinate system that is aligned with agent-heading direction. They determine the position of the center of the landmark relative to the agent. The parameter Γ determines the orientation of the surface on which the landmark is located relative to agent heading direction.

The above parameters can be determined from a single image given the camera parameters, the size of the landmark, and its height relative to the camera, based on the method proposed in [8] and further developed in [9]. The parameters are determined using the following expressions:

$$r = \sqrt{h^2\left([1+2\cot^2(\alpha_{hp})] \pm \sqrt{[1+2\cot^2(\alpha_{hp})]^2 - \frac{h_r^2+4z^2\cot^2(\alpha_{hp})}{h_r^2}}\right)} \quad (1)$$

$$R = \sqrt{r^2 - z^2} \quad (2)$$

$$\theta = \delta_{cam} - \alpha_{wc} \quad (3)$$

$$\gamma = \arccos\left(\frac{(r^2 - w_r^2)\tan(\alpha_{wp})}{2rw_r}\right) \quad (4)$$

$$\Gamma = \gamma + \theta + 90 \quad (5)$$

Table 1. Variables used in the equations of the localization algorithm and in their derivations in the Appendix, together with the figures in which they are illustrated. Input parameters describe the agent (e.g., focal length, f), or the constraints of the environment (e.g., the actual size of the landmark, w_r, h_r); all others are computed from the visual input to the localization module

Variable	Description	See Figure
w_r, h_r	Half of actual size (width/height) of bars of landmark (input: w_r = 126 mm, h_r = 86 mm)	17, 19
w_p, h_p	Perceived size (width/height) of bars of landmark (computed as the size of the bars detected on the CCD chip)	18
w_d, h_d	Distance between leftmost/topmost pixel of bar and image center (computed)	18
w_c, h_c	Distance between center pixel of bar and image center (computed)	
α_{wp}, α_{hp}	Angle subtended by w_p, h_p, respectively {computed: $\alpha_{hp} = \arctan\left[f h_p / \left(f^2 + h_d(h_d + h_p)\right)\right]$}	17, 18
α_{wc}, α_{hc}	Angle subtended by w_c, h_c, respectively	
δ_{cam}	Camera pan relative to agent heading direction (read from the turn-table)	
L	Focal point of camera	17, 18, 19
f	Camera focal length (input: f = 17.3 mm)	19
z	Relative height (height of camera minus height of center of landmark; input: z = 172.7 mm)	19
r	Distance between camera and center of landmark (Eq. 1)	2, 17, 19
R	Distance between camera and landmark horizontally (output: Eq. 2)	2
γ	Angle between normal of landmark plane and direction of landmark center relative to the camera horizontally (Eq. 4)	2, 17
γ_h	Angle between normal of landmark plane and direction of landmark center relative to the camera vertically	19
Γ	Angle between agent heading and plane containing the landmark (output: Eq. 5)	2
θ	Viewing angle of landmark center from agent (output: Eq. 3)	2

These equations are derived in the Appendix. Variable definitions are given in Table 1, and illustrated in the Appendix. There are two ambiguities in these equations. First, Eq. 1 has two solutions. This ambiguity can be resolved using the fact that the perceived size of the landmark is proportional to the distance. Second, Eq. 4 does not determine the sign of the angle γ. To determine it we rely on the distortion due to the perspective projection. We exploit the fact that the size of the vertical bars and that of the binary code are the same; therefore whichever is closer to the camera is perceived as larger.

2.3 Environmental influence on landmark based position estimation

The output of the localization algorithm is computed using Eqs. 1–5. Thus the variables in Table 1 determine the output of the algorithm. Some of these variables describe the agent (f, δ_{cam} and partly z, i.e., the height of the camera) and are determined in advance by calibration. Other variables describe the aspects of the environment, i.e., the landmark (h_r, w_r, and partly z, i.e., the height of the landmark from the ground). We might have *a priori* knowledge about these values or, alternatively, a stereo vision algorithm can be used to extract appropriate features from the scene and determine appropriate values.

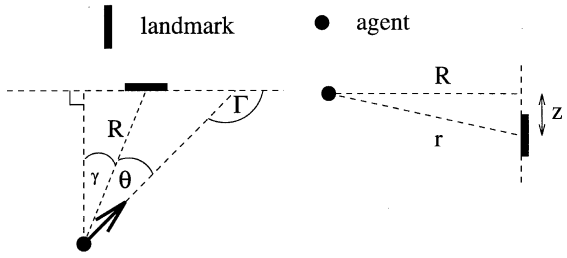
The values of the remaining variables (h_p, h_d, w_p, w_d, α_{wc}) are provided by the landmark recognition algorithm, and they are affected by the agent-environment interaction. Factors such as lighting conditions, distance from the landmark, viewing angle, and agent motion all contribute to determine the values of these variables, that is, they determine system performance. These factors are called task-level parameters [10]. Previous work carried out a study of the robustness of the algorithm [11]. In the next section we analyze the sensitivity of the localization algorithm to the values of the variables. In Sect. 4 we empirically determine the algorithm's performance as we vary the task level parameters.

3 Sensitivity analysis of the localization algorithm

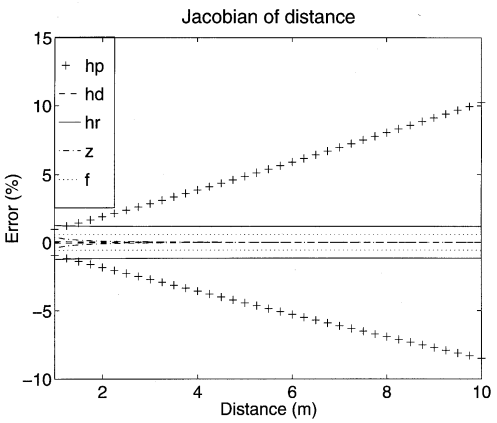
This section takes a theoretical approach to characterizing the behavior of the localization algorithm described by Eqs. 1–5. The advantage of selecting such a simple task as the landmark based ego-position estimation is that one can analyze and characterize the sensitivity of the algorithm to variations in input values. This helps to determine what information is most critical to performance, and how this information influences the behavior of the system. There are two reasons for performing such a sensitivity analysis. First, it helps to determine which task-level parameters need to be tested empirically because it demonstrates variables to which the algorithm is most sensitive. Second, it predicts the behavior of the system by estimating error bounds and predicting error biases [12–14].

The values of certain variables change depending upon the agent-environment interaction. For example, perceived size of the landmark, h_p and w_p, depends upon distance from the landmark, among other things. Thus if the algorithm turns out to be very sensitive to perceived height, it makes sense to design an experiment that manipulates the task-level parameter *distance* as an independent variable. Based on the sensitivity of such variables we can also provide error

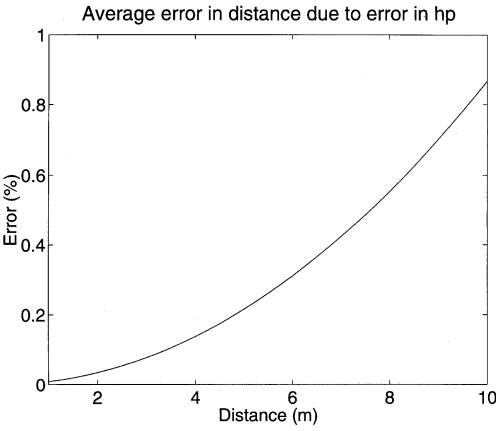
1



2a 2b



3a



3b

Fig. 1. The landmark: vertical bars followed by a binary code. The bars are constant to each landmark, serving as a header, while the binary code is a unique identifier

Fig. 2. Meaning of the parameters $(r, R, \theta, \gamma, \Gamma)$ determined by the localization algorithm **a** Top view of the setup. *Arrow* denotes heading direction of the agent. **b** Side view of the setup

Fig. 3. a Jacobian of distance measurement: the sensitivity of the distance measurement with respect to its five parameters as a function of distance. The curves predict bounds and biases on the error resulting from unit errors in the different variables. **b** The average of the error bounds due to error in the perceived height (h_p), predicting overestimation of R that increases with distance. (Error is percentage of the distance)

bounds, and assuming that errors are normally distributed we can determine expected mean error (bias). Other values in the algorithm are constant, determined by the physical device (focal length, f) or by constraints imposed on the environment (size of the landmark, $h_r,\ w_r$). Error in such parameters (e.g., calibration errors) lead to error bias.

To analyze sensitivity to the above parameters we computed the first derivative, i.e., the Jacobian of the system equations, or approximated the derivative by differences. The following sections describe the results of a sensitivity analysis for the estimate of distance, R, and then for the estimate of angle, Γ.

3.1 Analyzing the distance measurement

Equation 1 describes the distance between the focal point of the camera and the center of the landmark. Thus distance estimation, R, depends upon the variables of Eq. 1, most importantly perceived size (height), h_p, and vertical displacement, h_d, of the landmark. These values determine the size and position of the perspective projection of the landmark's vertical axis of symmetry and are provided by the landmark recognition algorithm. Since the vertical angle between the image plane and the plane of the landmark is constant from all viewing angles, these values depend only upon the distance, not upon the viewing angle. Distance estimation also depends upon the constants f, focal length, z, relative height, and h_r, actual landmark size (height).

Differentiating Eq. 1 with respect to these variables determines the sensitivity of the distance measurement. To obtain results in units comparable to empirical data, we computed differences instead of the derivatives. We calculated the effect of adding or subtracting one unit to each variable or constant. We defined one unit as 1 pixel for h_p and h_d, 1 mm for h_r, 10 mm for z and 0.1 mm for f. These differences are plotted in Fig. 3a, showing error in distance estimate (R) as a percentage of actual distance. The curves corresponding to h_p and h_d provide bounds on distance estimation error because these bounds are determined by agent-environment interaction (assuming 1-pixel errors). If we assume normally distributed errors, we can also predict bias in the error estimate by taking the average of the upper and lower curves. The curves for f, z, and h_r predict the nearly constant bias that would result from miscalibration of these constant values.

Figure 3a shows that distance estimation is most sensitive to h_p, the perceived size (height) of the landmark. Even a 1-pixel error can result in a significant error in the distance estimation. Notice that the upper and lower error bounds are not symmetric: the underestimation of the distance due to adding a pixel is smaller than the overestimation resulting from removing a pixel. Observing that in Eq. 1 estimated distance, R, is inversely proportional to perceived height ($r = ah_p^{-1} + c$), adding a pixel results in an underestimation error $err_{under} = \left[ah_p^{-1} + c\right] - \left[a\left(h_p + 1\right)^{-1} + c\right] = a\left[h_p(h_p + 1)\right]^{-1}$, while removing a pixel gives an overestimation error $err_{over} = \left[ah_p^{-1} + c\right] - \left[a\left(h_p - 1\right)^{-1} + c\right] = -a\left[h_p(h_p - 1)\right]^{-1}$, and $abs(err_{over}) > abs(err_{under})$. As-

suming normally distributed errors, and taking the average of the bounding curves, this predicts a positive error bias, a slight overestimation of R that increases with actual distance, as shown in Fig. 3b.

Referring again to Fig. 3a, we see that miscalibration of h_r or f introduces a nearly constant error bias. For example, using a smaller value of h_r is equivalent to placing the landmark further away from the camera, and so forth. One can also see that h_d and z have a negligible effect on the estimation.

3.2 Analyzing the angle measurement

The localization algorithm determines the position of the agent with respect to the landmark in terms of two angles, θ and Γ, computed by Eqs. 3 and 5. Assuming that the landmark is always centered in the image (i.e., α_{wc} = 0 in Eq. 3), θ is equal to the camera pan angle relative to the agent (δ_{cam}). Thus its accuracy depends solely on determining the pan angle.

The angle γ (Eq. 4), the angle between the camera and the normal of the plane of the landmark, depends on the perceived size of the landmark ($w_p,\ h_p$), and its perceived displacement (w_d, h_d), provided by the landmark recognition algorithm. As explained in Sect. 3.1, h_p and h_d depend solely on the distance. On the other hand, w_p and w_d, the size and position of the perspective projection of the horizontal axis of symmetry of the landmark, also depend on the viewing angle. This is because the horizontal angle between the image plane and the plane of the landmark is not constant, but corresponds to the viewing angle.

Similarly to the distance computation, the angle estimate is most sensitive to the perceived size of the landmark ($w_p,\ h_p$, shown in Fig. 4). Sensitivity depends upon both the distance and the angle. Distance is not very significant, but the viewing angle is of great importance. Surprisingly, sensitivity is greatest close to the right angle, at 90°, because the derivative tends to infinity. Although intuitively one might expect that this measurement would be the least sensitive when directly facing the landmark, Eq. 4 immediately explains the result: the derivative of the arccos(x) function tends to infinity if $x \to 1$, i.e., around 0°, which using Eq. 5 corresponds to the right angle. Computing differences instead of the derivatives causes the parameter of the arccos(·) function to exceed 1. In the algorithm such values are truncated to 1, so the numerical error remains bounded.

The estimate of Γ is also sensitive to the actual size of the landmark ($w_r,\ h_r$), its camera-relative height (z), and the camera focal length (f). These parameters are constant to the algorithm, and the effect of varying these parameters, including the focal length, is usually negligible (see Fig. 5).

3.3 Predictions based on the analysis

The result of the sensitivity analysis for the equations describing the localization algorithm can be summarized as follows. The most important variables are those that describe perceived size of the landmark, h_p and w_p. These values depend upon aspects of the agent-environment interaction (i.e., task-level parameters) such as actual distance, viewing angle, lighting conditions and motion of the agent. We also determined bounds on the error of the localization algorithm due to these parameters. For distance, R, error was independent of viewing angle. We also predicted a slight tendency to overestimate R (a positive error bias) as the actual distance from the landmark increases. For angle, Γ, error was found to be both distance and angle dependent, with greater sensitivity near 90° and increasing error at larger distances. Error in the focal length, f, and the actual height of the landmark, h_r, could cause constant biases, while other parameters have negligible effects.

Table 2. Illumination (in millilamberts) of the black and white regions of the landmarks, the wall and the floor. The different conditions correspond to different lights being turned on and off. Intervals are given, since the illumination varies with the viewing angle

Condition	Black region	White region	Wall	Floor
Full	1.1–1.3	16.0–18.7	14.5–17.4	5.4–6.6
Partial	1.0–1.2	14.2–16.1	12.5–15.3	4.6–5.6

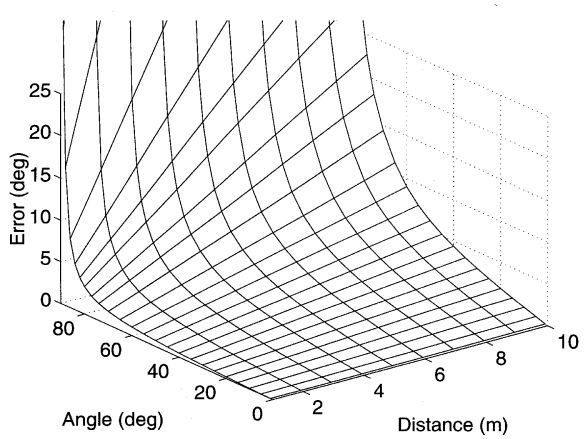
4 The experiment

The sensitivity analysis in the previous section investigated the effect of slight perturbations of system variables on localization behavior. Such perturbations may be the result of interaction with the environment or the result of miscalibration. The analysis provided predictions regarding what aspects of the agent-environment interaction should have the greatest influence on performance of the localization algorithm. In addition, it predicted the behavior of the algorithm in terms of error bounds and error biases. This section describes an experiment designed to test the performance of the algorithm in various conditions. We show that overall the behavior of the system fulfills some predictions, yet deviates from these predictions in interesting and meaningful ways.

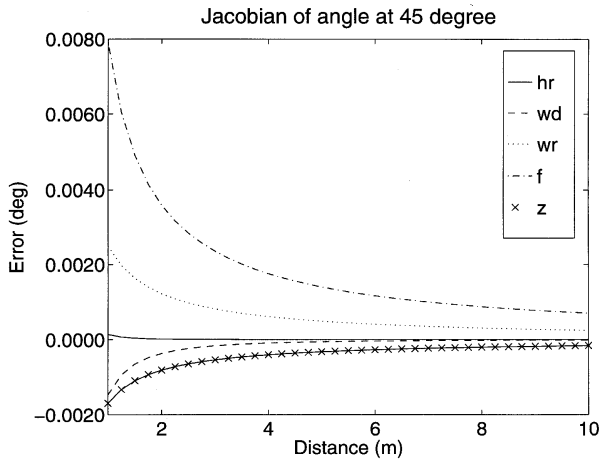
The goal of our design was to investigate the effect of manipulating task-level parameters, which describe the agent-environment interaction, on the behavior of the agent. Specifically, we investigated the effect of agent motion, actual distance, viewing angle, lighting conditions, and choice of thresholding strategy. We kept constant such parameters as landmark size and height, as well as focal length, height, and tilt of the camera.

4.1 Equipment

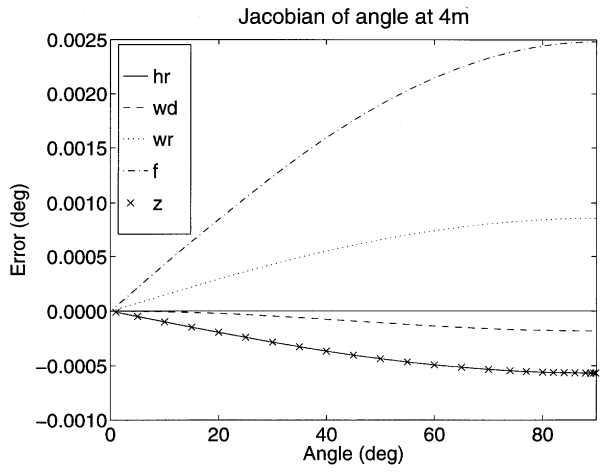
We used a TRC Labmate robot (Fig. 6). The 17.3 mm Cosmicar 1:1.4 lens with a Sony XC-77 CCD camera was mounted on a turntable. We used a SunVideo digitizer with double buffering capability allowing the parallelization of the grabbing and the processing operations. Our landmarks were printed in black on white paper. Their illumination in various conditions is summarized in Table 2, together with the data measured on the wall and on the floor. The different conditions result from turning on and off lights in the lab.



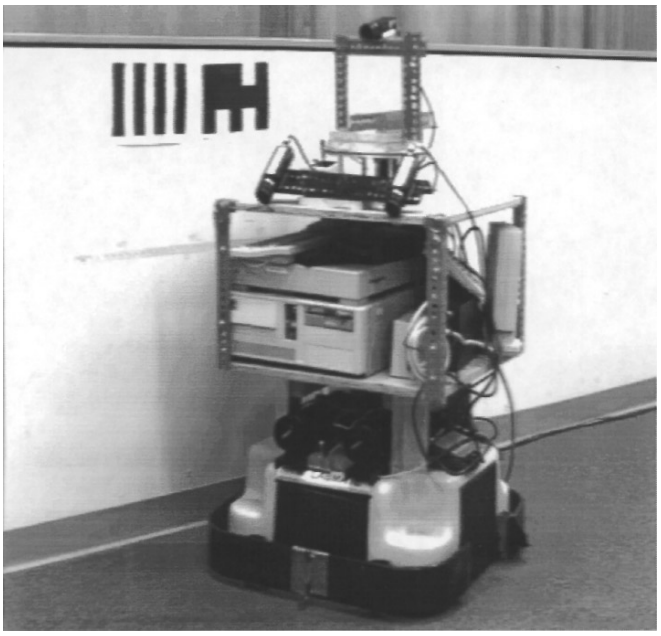
4



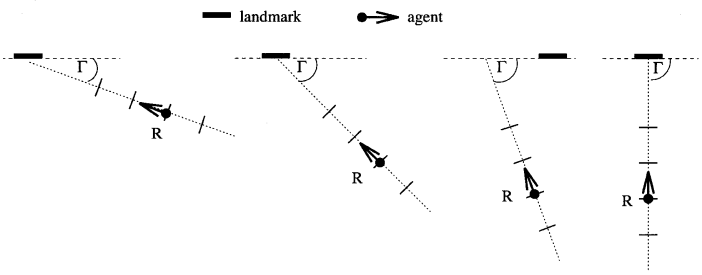
5a



5b



6



7

Fig. 4. Sensitivity of Γ angle measurement with respect to the perceived height, h_p. Sensitivity to the perceived width, w_p, is similar, with opposite sign

Fig. 5a,b. Jacobian of γ angle measurement. The sensitivity of the distance measurement with respect to five parameters as a function of distance at 45° (**a**) and as a function of angle at 4 m (**b**). Cross-sections of the 3D sensitivity plots are shown

Fig. 6. The TRC Labmate robot used throughout the experiments, and a landmark on the wall. The top camera on the turntable was used

Fig. 7. Top view of the four experimental trails corresponding to the Γ angles of 26°, 45°, 70°, and 90°. The agent moves along the *dotted line*, continuously maintaining the Γ angle relative to the plane of the landmark (*dashed line*). Measurements are taken at meter distances on the path (ticks)

4.2 Algorithm

The algorithm first grabs a 640 × 480 image and downsamples it by 2, taking every second pixel both horizontally and vertically. This step is followed by the gray-scale to black-and-white conversion, which used either a constant predefined threshold value of 128 or adaptive thresholding using the average pixel intensity of the whole image. The next step is to search for and find the landmark. The perceived size of the landmark was to be extracted by finding the corners of the landmark and determining the height and width as the difference of these data, achieving accuracy of one-half pixel. Using these data, Eqs. 1 and 4 compute the parameters relative to the camera center, and these values are transformed to the agent relative values using the δ_{cam} parameter (the angle between the camera and the heading direction of

the agent). The camera is mounted on a turntable and the value of δ_{cam} is provided by the odometry of the turntable. According to our measurements, this odometry reading is very reliable, it remained within $\pm 0.5^\circ$ even after very long test sequences.

The algorithm ran on a SPARCstation 2. One cycle of the algorithm, that is, grabbing an image, converting from gray-scale, recognizing the landmark, and determining the camera relative position of the landmark, took 0.28 s and 0.31 s for the constant and adaptive thresholding, respectively, when not using the parallel feature of the frame grabber, and 0.21 s and 0.24 s, respectively, when grabbing and processing were performed in parallel.

4.3 Experimental design

The experimental design was a $4 \times 4 \times 2 \times 2 \times 2$ factorial design with repeated measures. The independent variables corresponded to task-level parameters that describe the agent-environment interaction, specifically: actual distance, viewing angle, agent motion, and lighting conditions. Thresholding strategy provided the fifth independent variable.

The specifics of the design reflected a combination of theoretical and practical considerations. First, we took measurements at four distances (R = 2, 3, 4, and 5 m). Distance was sampled linearly because the effects that were predicted to be most important by our analysis were nearly linear in distance. The range of distances was limited by the size of our laboratory on the one hand, and by the fact that camera blur becomes a strong effect at distances closer than 2 m. Next, we used four levels of viewing angle (Γ = 26°, 45°, 70°, and 90°). Viewing angles were selected so that all different regions of the predicted error curve were sampled. Smaller angles were constrained by the size of the agent and the minimum distance. Two levels of motion (at rest and in motion) were used to demonstrate qualitatively that motion is an important factor, without determining the differential effect of velocity. Measurements were taken using two different types of threshold (constant and adaptive). Our main goal here was to show that internal properties of the algorithm influence the results not merely in a binary fashion (e.g., in dim light adaptive thresholding must be used), rather, as we shall see, threshold type makes a difference even in bright light. Finally, two levels of lighting (full and partial) were used. The number of lighting levels were constrained by the fact that constant thresholding did not produce reliable landmark recognition at dimmer lighting levels.

Two dependent measures were taken at each point, error in estimated distance (as percentage of actual distance) and error in estimated angle (degrees). We repeated each performance measurement four times. The number of measurements required can be estimated in advance by estimating both the size of each predicted effect and the expected error variance [15]. The actual number that we estimated was considerably larger than four, and carrying out these experiments is very expensive in terms of time. Thus in order to minimize resource utilzation we measured actual effect size and error variance as the experiment proceeded using a post hoc power analysis, reported below, to determine when enough measurements had been taken. We were therefore able to detect most of the phenomena of interest without consuming an inordinate amount of resources.

4.4 Method and procedure

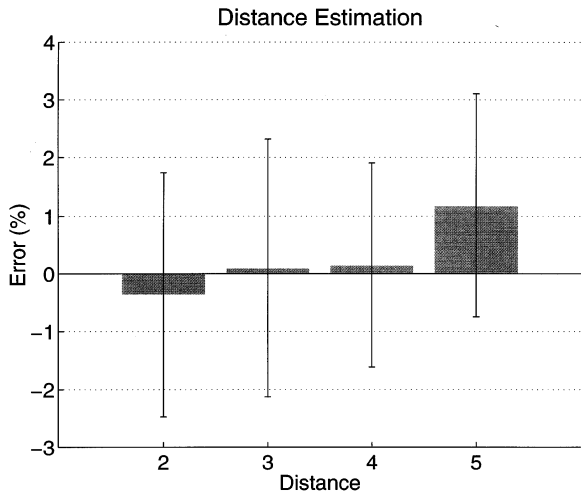
Measurements were taken at four different distances, R, and 4 different Γ angles, as illustrated in Fig. 7, both when the agent was at rest, and when it was in motion, in two different lighting conditions, and using two different gray-scale to black-and-white thresholdings. Each angle corresponded to a path: one set of measurements was taken moving along a path, moving the agent either away from or toward the landmark while maintaining a constant angle. To ensure a straight-line motion, before each run the agent was positioned at either end of the path at the required angle, with the caster wheels aligned. At each position the image was stored and processed off-line with the two different thresholding algorithms. Actual distance was measured manually using a metal measuring tape. Viewing angles were measured using triangulation. For each type of data, that is, at a given distance, angle, light, thresholding, motion combination, we made four measurements.

For the at-rest measurements the agent was moved along each path to the required distance, and two images were stored, one under each lighting condition. The agent was then moved to the next distance, and the images were recorded. The agent was run four times along each path. Recordings were also made while the agent was moving with a constant speed of 100 mm/sec. In this case the agent was positioned at either end of the path with its caster wheels aligned, but outside of the measurement area, for example 1.5 m from the landmark. It began moving, storing the images at each meter. For this we relied on the odometry readings of the agent, corrected by manual measurement of the actual distance traveled. The agent was run eight times along each path, four times per lighting condition.

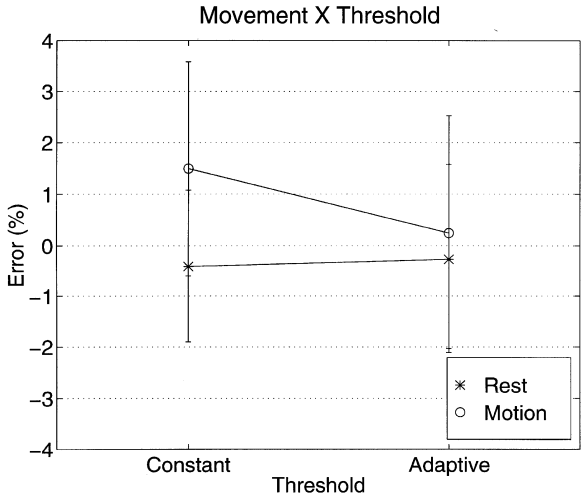
4.5 Results

To determine the systematic effects of environmental manipulations on the agent's estimates of relative location, a first analysis of variance (ANOVA) [15] was performed on measured errors in distance estimates, R and a second ANOVA was performed on the measured error in angle estimates, Γ. Analysis of variance is typically performed to control type I error, or α, the probability of reporting an effect as significant when none actually exists. Below we report the F ratio (roughly, the ratio of effects due to manipulation of the independent variables to the experimental error). F ratios near 1 imply no effect of the manipulation whereas those greater than 1 imply some effect. We also report the significance level of the F ratio, the probability, p, of type I error in comparison to a predetermined α level (e.g., α = 0.05, a typically reported criterion level for statistical significance). Associated with each F ratio are two degrees of freedom, describing the shape of the F distribution.

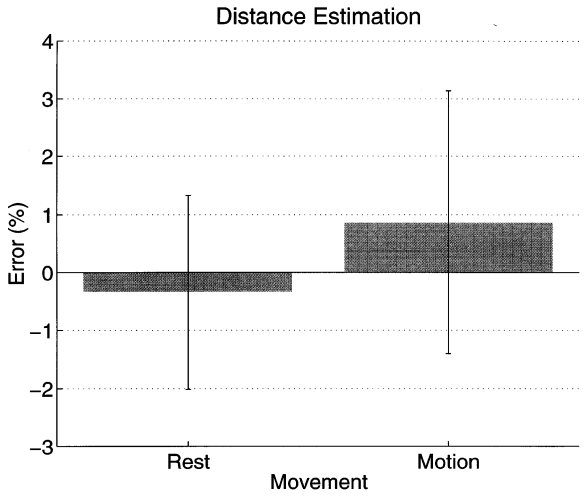
Next we performed a post-hoc power analysis [15] to determine the size of each effect and the *power* of our experiment to detect each effect. The power analysis allows



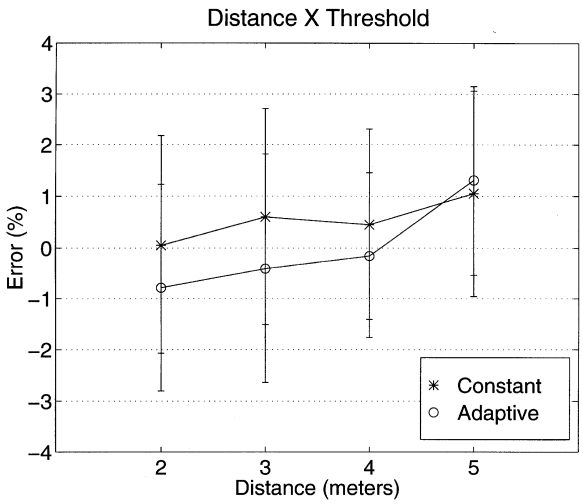
8a



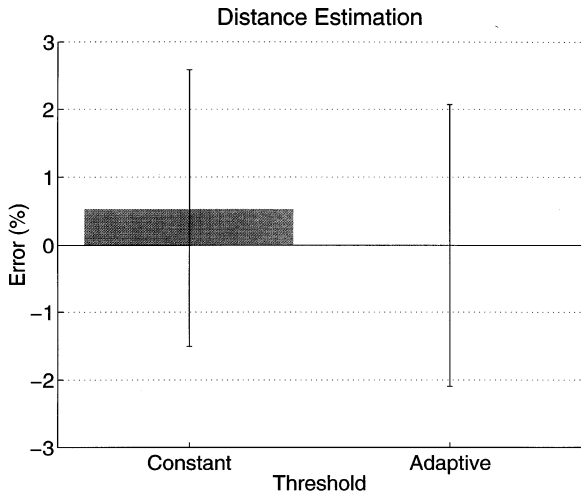
9a



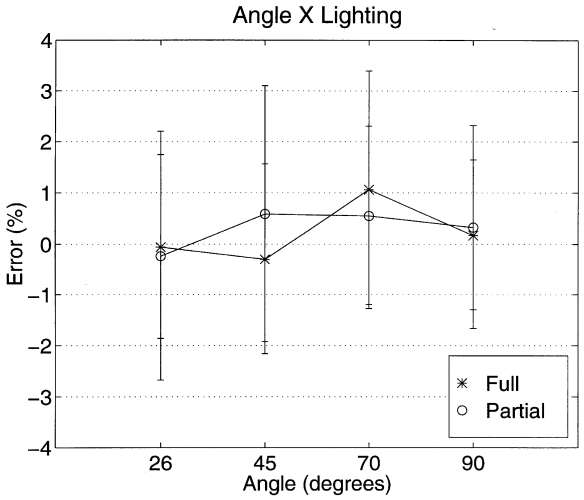
8b



9b



8c



9c

Fig. 8a–c. Main effects for distance estimation

Fig. 9a–c. Two-way interactions for distance estimation

us to also measure type II error, or β, the probability of reporting no effect when one actually exists. Thus, we report not only F ratios and their associated significance levels (α values) but also *effect size* (ω^2) and power ($1-\beta$). Effect size reflects the proportion of variance (due both to experimental manipulations and to inconsistency in agent performance) that can be attributed to differences in experimentally manipulated conditions; in our case it reflects the proportion of overall variance accounted for by manipulating the task-level parameters. Below, we report all effects for which $\omega^2 > 0.10$. These are either main effects (significant difference in performance caused by the manipulation of a single independent variable), or interactions (one of the independent variables does not have a constant effect at all levels of the other independent variable). We first discuss the effects of the experimental manipulations on estimation of distance R and then on estimation of angle Γ.

In both cases the data displayed marked heterogeneity of variance, thus we performed a Geisser-Greenhouse correction to determine statistical significance of the F ratios from both analyses. The Geisser-Greenhouse correction assumes maximal heterogeneity of variance [16]; this conservative approach seemed warranted given the large degree of heterogeneity observed.

Accuracy of distance estimation. The analysis turned up a number of significant effects which show the importance of agent-environment interaction in determining the ability of the agent to estimate distance from a landmark. Table 3 reports all effects for which $\omega^2 > 0.10$, i.e., those which account for at least 10% of the variance in their respective measurements. We first discuss effects that were found to be statistically significant, and then we discuss power for one effect that was found not to be significant.

The significant main effects are shown in Fig. 8. First, we found a main effect of distance ($F_{3,9} = 7.239, P < 0.10$). This confirms the predictions of the sensitivity analysis for a positive error bias that increases with distance. However, at the closest distance, we see a slight negative bias; this could reflect a constant bias introduced, for example, by a slight miscalibration of f or h_r. Next, there was an even stronger main effect of motion ($F_{1,3} = 12.078, P < 0.05$). In general, the agent estimated distance quite accurately when at rest, but overestimated distance when in motion. Finally, the strongest main effect was that of threshold (gray-scale conversion type; $F_{1,3} = 29.060, P < 0.05$). Not surprisingly, the agent was significantly more accurate in determining distance using the adaptive than the constant threshold.

Next, three two-way interactions were found to be significant, showing that a combination of factors can conspire to influence agent performance. The largest of these was a strong interaction between motion and threshold ($F_{1,3} = 53.358, P < 0.01$; Fig. 9a). This indicates that the agent estimates distance very accurately when at rest, and that the type of threshold helps very little if at all. However, adaptive thresholding helps significantly when the agent is in motion. Second, there was an interaction between distance and threshold ($F_{3,9} = 15.711, P < 0.05$; Fig. 9b). The adaptive threshold produces a negative bias when the agent is close to the landmark but more accurate results at intermediate distances. At a distance of 5 m, the two methods show little difference. This could be due to projection and illumination effects, so that at 5 m contrast is not helping. Third, we found an interaction between angle and lighting condition ($F_{3,9} = 7.371, P < 0.05$; Fig. 9c). This interaction is more difficult to interpret since the pattern of sensitivity differs depending upon the lighting, but it appears that under conditions of full lighting the system is more sensitive to oblique angles. This may be due to the angle dependence of the illumination and the interplay of the projection and the illumination effects.

Additionally, we found two large three-way interactions. Of these, the interaction among motion, lighting and threshold was the strongest ($F_{1,3} = 44.981, P < 0.01$; Fig. 10). Here lighting level can be seen to modulate the interaction observed between motion and threshold. Adaptive thresholding helps the agent more in partial than in full lighting conditions. The three-way interaction among motion, angle, and lighting was also quite strong ($F_{3,9} = 28.779, P < 0.01$; Fig. 11). One can see from Fig. 11 that motion modulates the basic two-way interaction (above) between angle and lighting. When at rest, in both full and partial lighting conditions, the agent is very accurate in estimating distances. When in motion, however, lighting makes a rather large difference in the agent's performance; thus the interaction between angle and lighting affects the agent mainly when in motion.

We also briefly address the issue of an effect that was not statistically significant, an effect of viewing angle on the agent's errors in estimating distance. Although this was not predicted by the theoretical model of the previous section, Table 3 reveals that the effect of angle is rather large ($\omega^2 = 0.3988$) and approaches significance ($F = 4.583, P < 0.25$); remember that our choice of the Geisser-Greenhouse correction for heterogeneity of variance is a very conservative one. Looking at the power for the distance effect (Table 3), one can see that given the number of observations (four) and the variance in these measurements, we had approximately a 76% chance of observing such an effect if it indeed exists. We estimate that detecting this effect would require several additional observations; for example, an additional five observations would virtually ensure detection of any effect of viewing angle ($1-\beta > 0.99$). Nevertheless, we decided that detection of this effect did not justify the additional resources required.

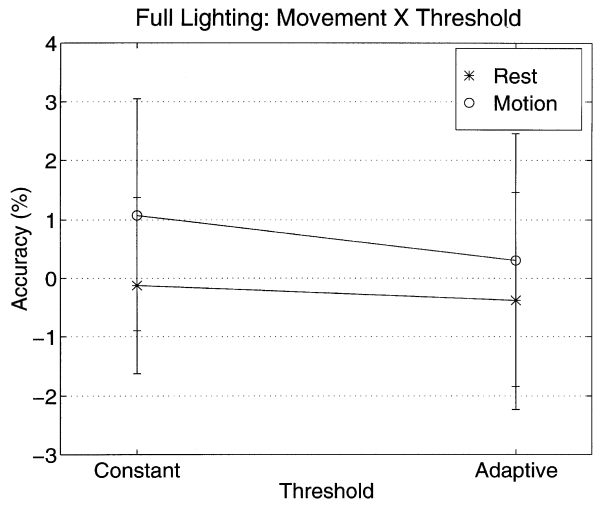
Finally, we consider how the agent fared in terms of performing within predicted error bounds. Figure 12 shows performance of the resting vs. moving agent, plotted against the bounds predicted under the assumption of a 1-pixel error. As we can see, the means all fall within the predicted bounds, and standard deviations fall more or less within the bounds as well. Deviations from the predictions occur at closer distances, however, implying errors greater than 1-pixel in calculating perceived height. In addition, the error bars also give an indication of the heterogeneity of variance that required the Geisser-Greenhouse correction to our ANOVA results.

Accuracy of angle estimation. This analysis turned up a number of significant effects which show the importance of agent-environment interaction in determining the ability of

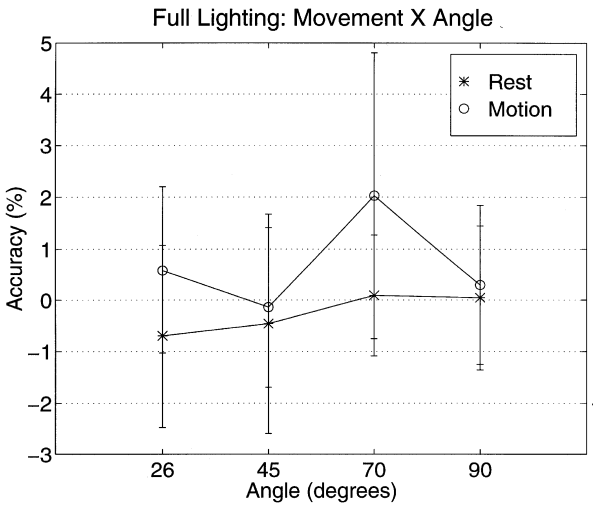
Table 3. Statistically significant and marginally significant effects for distance estimation after application of Geisser-Greenhouse correction for heterogeneity of variance. Reported are all effects whose size (ω^2) was greater than 0.10 as determined by a post-hoc power analysis

Effect	F ratio		Effect size (ω^2)	Power($1-\beta$)
Distance	7.239	*	0.5391	0.96
Movement	12.078	**	0.5807	0.98
Threshold	29.060	**	0.7781	>0.99
Movement × threshold	53.358	***	0.7659	>0.99
Distance × threshold	15.711	**	0.5797	0.98
Angle × lighting	7.371	*	0.3739	0.72
Move × light × threshold	44.981	***	0.4073	0.78
Move × angle × light	28.779	**	0.5656	0.98
Angle	4.538	+	0.3988	0.76
Angle × threshold	2.732	+	0.1397	<0.30
Angle × light × threshold	3.375	+	0.1002	<0.30

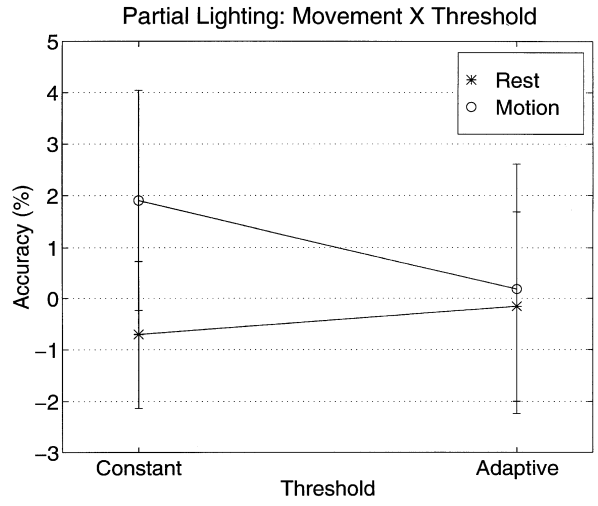
Significant effects: * $P < 0.10$; ** $P < 0.05$; *** $P < 0.01$. Marginally significant effects: + $P < 0.25$



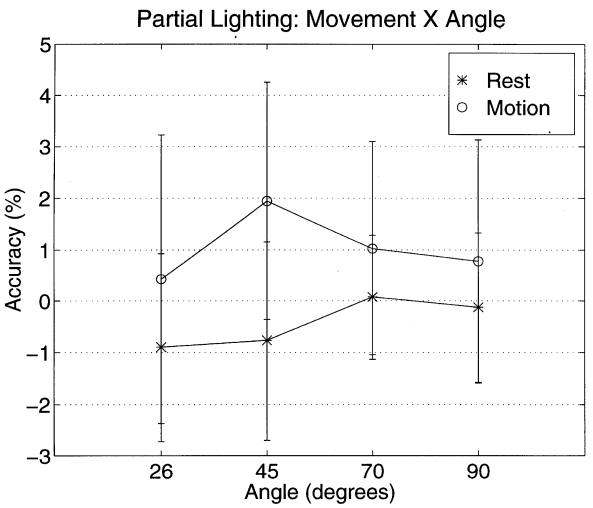
10a



11a



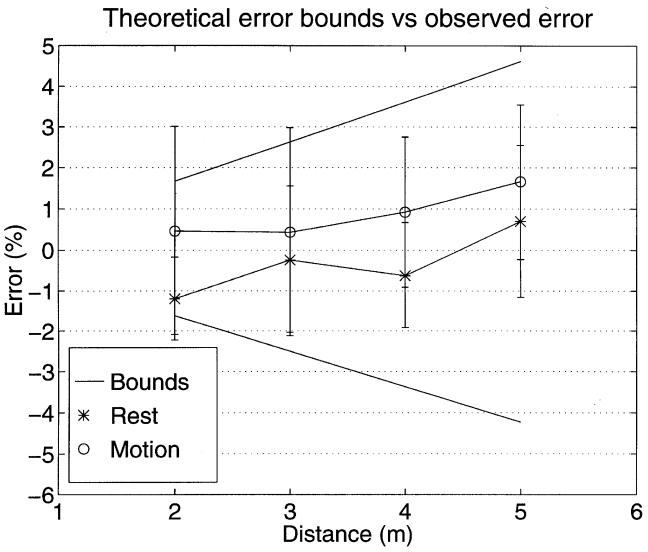
10b



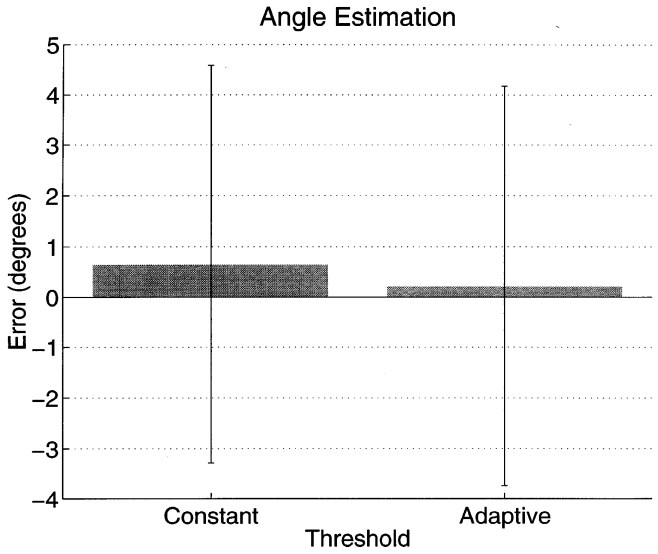
11b

Fig. 10. Three-way interaction for distance estimation: movement × lighting × threshold

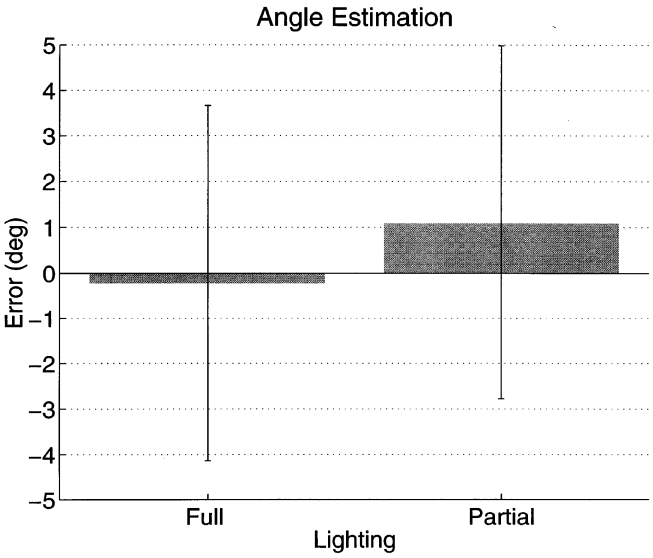
Fig. 11. Three-way interaction for distance estimation: movement × angle × lighting



12



13b



13a

Fig. 12. Theoretically determined error bound (due to 1 pixel error) on distance computation together with the measured data

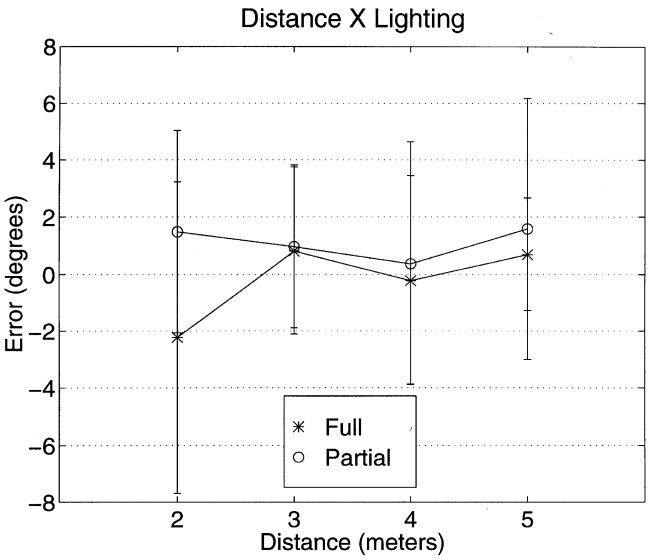
Fig. 13a,b. Main effects for angle estimation

the agent to estimate relative angle with the landmark. Table 4 reports all effects for which $\omega^2 > 0.10$, i.e., those which account for at least 10% of the variance in their respective measurements. Overall, the analysis of angle estimation accuracy turned up fewer and smaller effects than did the analysis of distance estimation errors, due mainly to higher variance in angle estimation errors. We first discuss effects that were found to be statistically significant, and then we discuss power for one effect that was not found to be significant.

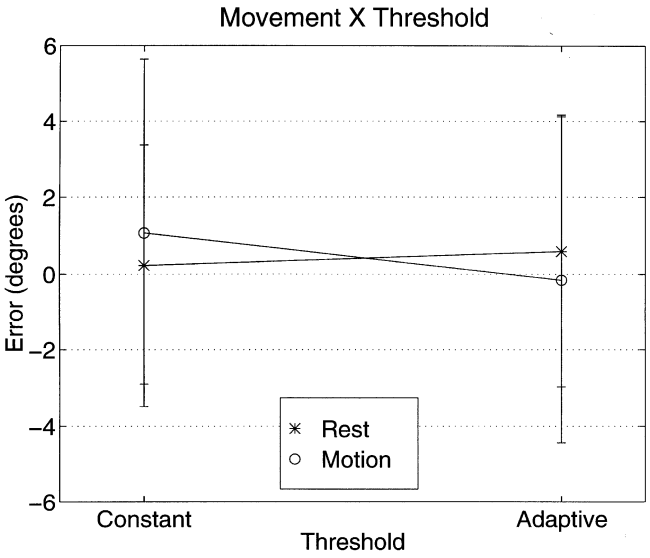
We found two main effects (Fig. 13). The first was a significant effect of lighting ($F_{1,3}$ = 37.695, $P < 0.05$). The agent was significantly more accurate in estimating angle in the full lighting condition. In addition, there was also a significant effect of gray-scale conversion algorithm ($F_{1,3}$ = 11.052, $P < 0.10$). As with distance estimation, the agent was more accurate in estimating pose angle using the adaptive thresholding algorithm.

In addition, we found five two-way interactions (Fig. 14). First, we found a significant interaction between distance and lighting ($F_{3,9}$ = 31.592, $P < 0.05$). Under full lighting conditions the agent is more accurate at estimating pose angle, although the pattern of biases is somewhat difficult to interpret. In partial lighting the agent is not as successful and actually improves with distance. Second, we found a significant interaction between lighting condition and movement ($F_{1,3}$ = 18.716, $P < 0.05$). When at rest the agent is quite accurate in estimating pose angle, whereas lighting conditions strongly affect the behavior of the moving agent. Third, we found a significant interaction between movement and distance ($F_{3,9}$ = 14.122, $P < 0.05$). This interaction shows differing patterns of sensitivity to distance depending upon whether the agent is in motion.

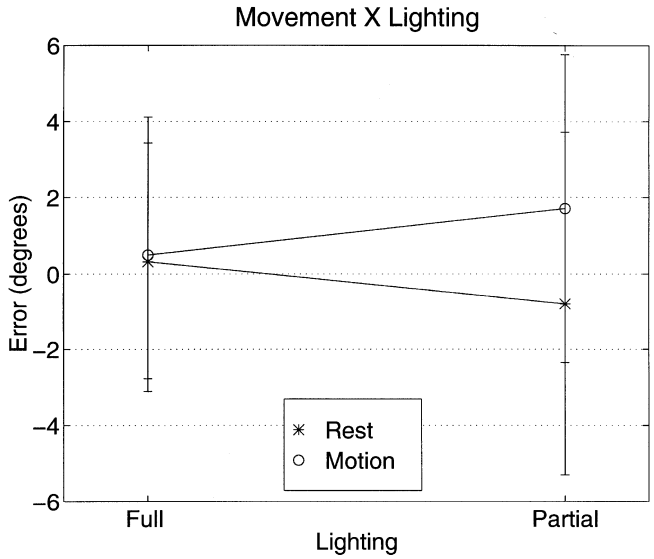
Fourth, we found a significant interaction between movement and threshold ($F_{1,3}$ = 12.067, $P < 0.05$). The adaptive threshold improves performance when the agent is moving but does not improve performance when the agent is at rest.



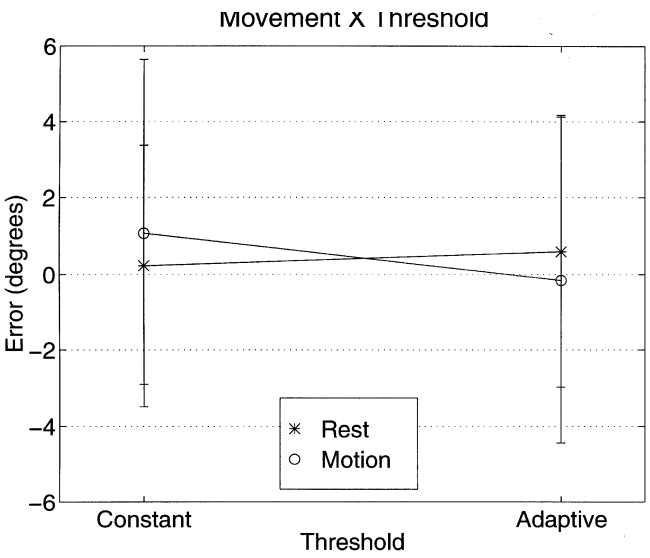
14a



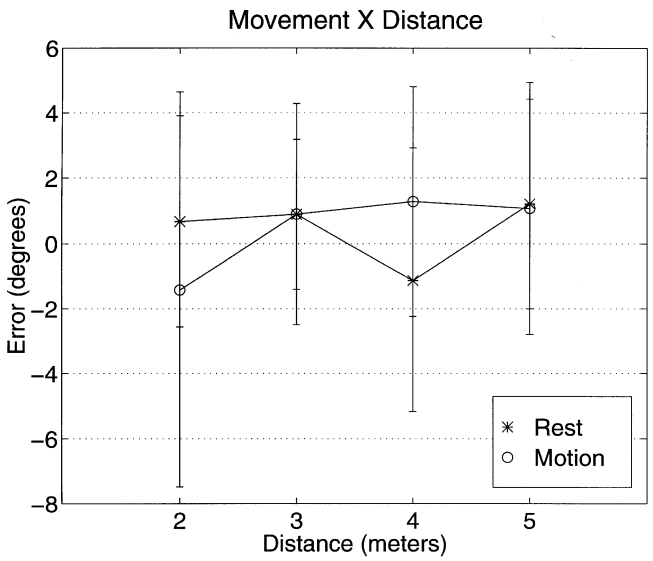
14d



14b



14e



14c

Fig. 14a–e. Two-way interactions for angle estimation

Table 4. Statistically significant and marginally significant effects for angle estimation after application of Geisser-Greenhouse correction for heterogeneity of variance. Reported are all effects whose size (ω^2) was greater than 0.10, as determined by a post-hoc power analysis

Effect	F ratio		Effect size (ω^2)	Power($1-\beta$)
Lighting	37.695	***	0.8210	>0.99
Threshold	11.052	**	0.5568	0.97
Distance × lighting	31.592	**	0.7415	>0.99
Movement × lighting	18.716	**	0.5254	0.95
Movement × distance	14.122	**	0.5516	0.96
Movement × threshold	12.067	**	0.4089	0.78
Angle × threshold	7.795	*	0.3891	0.76
Move × distance × light	9.274	*	0.2795	0.50
Distance	3.963	+	0.3571	0.68
Move × distance × angle	3.253	+	0.1368	<0.30
Distance × angle × light	3.355	+	0.1421	<0.30
Distance × angle × threshold	4.249	+	0.1860	<0.30

Significant effects: * $P < 0.10$; ** $P < 0.05$; *** $P < 0.01$. Marginally significant effects: + $P < 0.25$

Fifth, there was a small but significant interaction between angle and threshold ($F_{1,3} = 7.795, P < 0.10$). The constant threshold algorithm appears to be more sensitive to the viewing angle.

Finally, we discovered a three-way interaction among movement, distance, and lighting (Fig. 15). While the precise nature of the interaction is difficult to interpret, it seems to indicate that the agent's sensitivity to distance is most relevant when the agent is in motion and in partial lighting conditions.

Next, we discuss the effects not observed in this experiment. Most interestingly, significant main effects of neither distance nor viewing angle on estimation of Γ were observed. To be sure, the theoretical analysis did not predict any bias; nevertheless it did predict great deal of sensitivity to these parameters. A relatively large but nonsignificant effect of distance does appear in Table 4. However, the post hoc power analysis revealed that this experiment had only a 68% chance of detecting this effect. Several additional observations may have shown this effect to be significant; nine observations would have provided a 97% chance of detecting this effect. Distance was also a factor in two interactions, probably as a result of the algorithm's sensitivity to this parameter. Additionally, the significant angle × threshold interaction was probably due to the algorithm's sensitivity to viewing angle in addition to the overall large biasing effect of threshold. Thus, estimation of viewing angle reveals sensitivity to distance and viewing angle but little if any bias.

Finally, we consider performance with respect to predicted error bounds. Figure 16 shows performance at rest vs. in motion, plotted against the bounds by the sensitivity analysis. All means fall within the predicted bounds, although not all standard deviations respect predictions. The figure also shows marked heterogeneity of variance among different conditions.

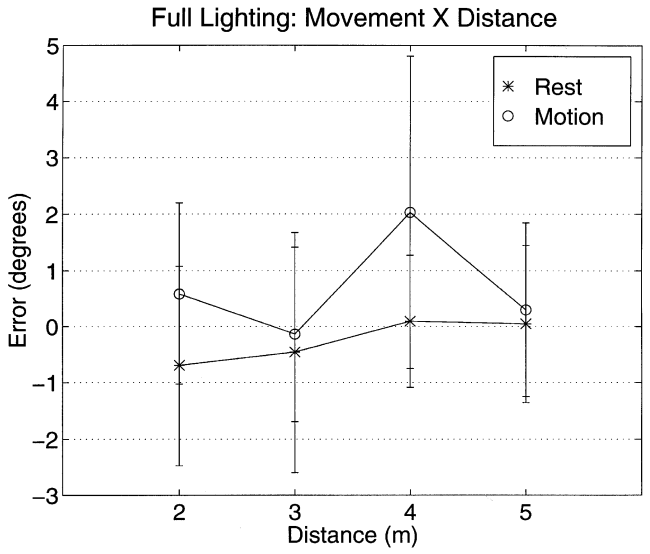
4.6 Discussion

Our empirical investigation uncovered a number of important facts about the performance of the agent in a realistic setting. First, we were able to verify the performance of the agent's vision algorithm. Comparison of the theoretical bounds on error performance with the standard deviations observed in the experiment showed that the agent does indeed perform within the predicted bounds. Although this may not seem surprising, it is quite important given the number and the complexity of effects that the agent-environment interaction was shown to have upon performance.

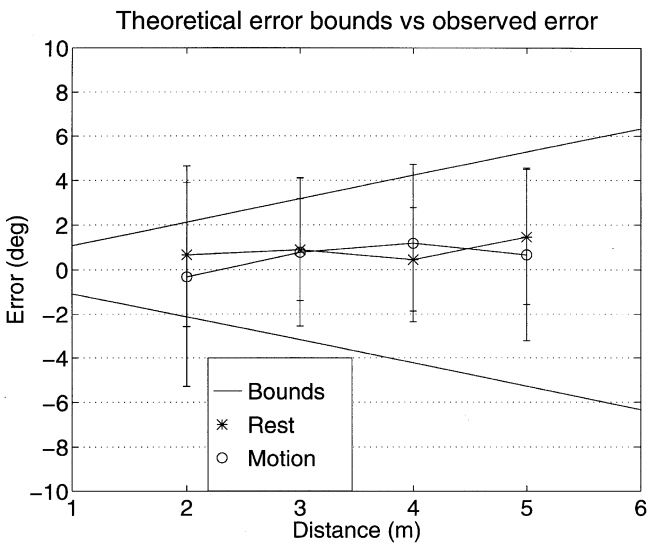
We found a main effect of actual distance on distance estimation. This matched our theoretical prediction of positive error bias in the estimation of distance. In addition, the importance of lighting level, both as a main effect in angle estimation and as a factor in statistical interactions in estimation of both parameters, is interesting especially since it was not addressed in the theoretical analysis. These results emphasize the basic soundness of this approach to combining theoretical analysis with empirical testing. They underscore the importance of empircal experimentation both to validate existing theoretical models and also to help create them.

The most critical aspect of agent-environment interaction to influence the performance of the vision algorithm was motion of the agent. The agent performed quite well when at rest but less well when in motion. This conclusion is supported by a significant main effect of motion, accompanied by increased variability, in distance estimation errors. We identified a significant overestimation of distance when the agent was in motion. In addition, motion participated in statistical interactions, for estimates of both distance and angle. Motion decreased performance when the going got tougher for the agent, in particular under conditions of degraded lighting, when angles were far from 90°, and when the less reliable constant thresholding algorithm was used. This finding is crucial because in typical behavioral situations either the agent, the landmark, or both are moving.

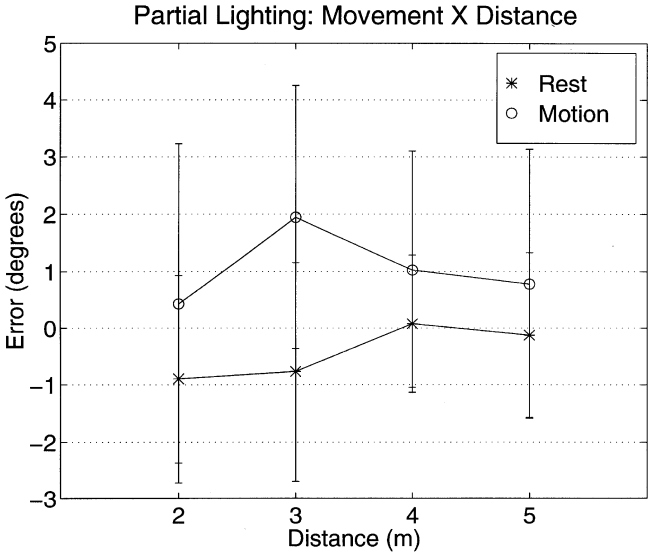
Another important factor in vision algorithm performance was thresholding strategy, constant versus adaptive. Perhaps not surprisingly, the adaptive thresholding algorithm consistently outperformed the constant threshold strategy, the latter introducing a significant bias in estimation of both parameters. In addition, this factor interacted statistically with environmental factors: the adaptive thresholding algorithm helped more under poorer illumination and when the agent was in motion.



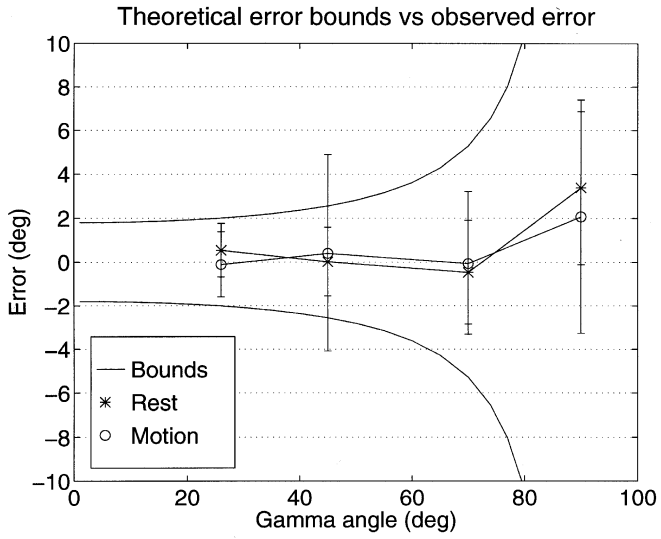
15a



16a



15b



16b

Fig. 15a,b. Three-way interaction for angle estimation: movement × distance × lighting

Fig. 16a,b. Theoretically determined error bound on angle computation together with the measured data. **a** Bound on measurements at 70° Γ angle. **b** Bound on measurements at $R = 5\,\mathrm{m}$ distance

Finally, and perhaps most importantly, we have shown that the variables that change during dynamic interaction with the environment produce effects that interact, conspiring to effect the performance of the agent in complex ways. Although our theoretical analysis hinted at the presence of these effects, it was not able to predict precisely what they would be. This points to the importance of actually performing empirical evaluations such as this. The complexity of agent-environment interaction can have systematic effect on the performance of the agent that may be quite subtle and difficult to predict, even given *a priori* knowledge of the construction of the autonomous agent.

5 Conclusions

The goal of this study was to put forward a methodology for prediction and verification of the performance of robotic agents that act in partially unknown and ever-changing environments. Robotic agents are not "black boxes"; we have known parametric models because they are synthetic. The only source of uncertainty comes from the physics of the component devices: dark current in the CCD camera, friction in gears, and so on. Fortunately, such effects tend to be small, and we understand well how to model them. A far more important source of variability, however, arises from the interaction of the agent with an unknown and ever-changing environment. As we have shown, some aspects of this interaction have systematic effects on agent performance, other aspects are best modeled as stochastic processes. Our methodology

consists both of theoretical analysis and empirical experiment. We address the basic methodological question of how well we can model the agent-environment interaction based on physics and geometry, what factors are best explored by empirical experimentation, and how analysis and experimenation can interact. It is this conceptual separation between agent and environment, along with an analysis of their interaction, that we believe is our contribution.

We selected a relatively constrained yet nontrivial task: vision-based localization of a robotic agent. We modeled the agent in terms of its visual capabilities, such as position and orientation of its camera system, overall height, speed of movement, and so forth. The environment is modeled as a set of landmarks with known visual/geometric characteristics. The interaction of the agent vis-a-vis its environment is characterized by the parameters R and Γ that describe the location of the agent relative to the landmarks. We derive analytic functions describing the interaction because of the constrained nature of the task. In our analysis of performance we restricted ourselves to task-level parameters, assuming successful landmark detection.

Questions in such empirical evaluation include: how many experimental conditions must one consider, and how many observations must one perform in order to validate the model? We have seen that in the robotic agent context the number of experimental conditions is determined by the number of task-level parameters and combinations of parameters. These limits serve to constrain the task level parameter space, and the experiments can be viewed as a search in this constraint space. However, it is the analysis of variance (Sect. 4) that reveals how many of the task-level parameters are dependent or independent. The dependency increases the dimensionality (one must consider their combinations) of the task-level parameter space. Thus it may be necessary to consider complex factorial designs as we present above.

In theory the number of required observations may be estimated using the error bounds and biases predicted by a sensitivity analysis. However, the actual spread (standard deviation) of the empirical error distributions is the real determinant governing the number of observations required. In addition, the cost of performing each observation is an important factor in determining how many observations to take, especially when running a complex factorial design. Such considerations point to a predict-experiment loop to determine the true nature of the agent-environment interaction, in which the post hoc power analysis plays an important role. In our experiment we performed the number of observations necessary to unambiguously detect most of the phenomena of interest, without consuming an inordinate amount of resources.

The results obtained in the analysis of variance led to some expected and some unexpected conclusions. For example, we anticipated that adaptive thresholding would improve overall performance somewhat and even more under conditions of partial illumination. We did not necessarily anticipate that adaptive thresholding would improve the performance of the agent so much more when in motion than when the agent was at rest.

The results of the analysis regarding the effects of motion were in general more difficult to predict. Although we anticipated that a somewhat larger error variance would result from motion, it was rather surprising to observe large systematic effects of motion on agent behavior. It is important to point out that these effects were not anticipated by the sensitivity analysis, indeed theoretically the agent should perform similarly in the two cases.

In addition, the sensitivity analysis did not predict the dependencies among task-level parameters. In the empirical analysis, however, these dependencies showed up as statistical interactions, giving some insight into how task-level parameters combine to determine performance. In principle, predictions of such dependencies may be difficult to make *a priori* because purely analytical tools are not always sophisticated enough to account for the complexity of the agent-environment interaction.

The methodology can be summarized as a cycle [17], first determining the parameters of the system (parameterization), then making predictions based on a model of the system (forward problem), and finally observing actual behavior in experiments and verifying, or updating the model based on these results (inverse problem). These considerations point to the need for empirical performance analysis, not simply to verify models of performance, but to help design them. Hence experimentation is crucial to complete analysis of any robotic system.

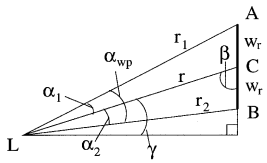
Ultimately we expect such performance analyses to lead to the development of self diagnostic procedures for agents that act in partially known, dynamic environments. Forstner [18] has advocated the development of such self-diagnostic programs, which he calls “traffic light programs”. A traffic light program signals “green” if the result is correct, “yellow” if it may be correct but needs checking, and “red” if it is incorrect or no result is produced. Examples of factors that might cause “yellow” signals in localization are lighting level, agent speed, and distance from the landmark. Knowing the limitations of the agent by exploring the parameters and interactions which have the most significant influence on behavior will allow us to develop such procedures.

In the limit, we envision that a fully developed understanding of such performance analysis methodologies will enable the construction of autonomous agents that are able to carry out sophisticated analyses of their interactions with unknown environments. Such a system would be able to develop knowledge of significant dependencies and alert itself when to be more or less cautious in its environment. Such systems would be truly flexible and robust to complex sets of changes in environmental situations. The current study is but a first step toward this goal.

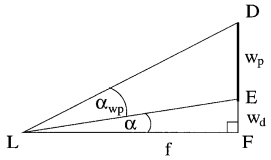
Acknowledgements. We thank Harris Romanoff for his tremendous help in performing the experiments and Henrik I. Christensen for his insightful comments. The work was supported by Army grant DAAH04-96-1-0007, DARPA grants N00014-92-J-1647 and DAAH04-95-1-0067, and NSF grants IRI93-07126 and SBR89-20230.

Appendix

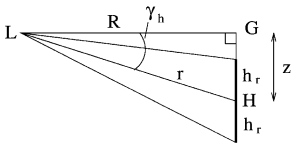
Here we derive Eqs. 1–5, used to compute the position of the landmark relative to the agent.



17



18



19

Fig. 17. The top view of the setup: the plane crosses the camera center and the center line of the bar. *Thick line* denotes the bar

Fig. 18. The image of the landmark on the CCD array (top view)

Fig. 19. The view of the setup from the side. *Thick line* represents the landmark

Figure 17 shows the top view of the setup (the meaning of the variables is explained in Table 1):

Law of cosine for $\triangle_{ACL}$:

$$r^2 + w_r^2 - 2rw_r \cos(180 - \beta) = r_1^2 \quad \text{(A-1)}$$

Law of cosine for $\triangle_{BCL}$:

$$r^2 + w_r^2 - 2rw_r \cos\beta = r_2^2 \quad \text{(A-2)}$$

$$(\text{A}-1) + (\text{A}-2) : 2\left(r^2 + w_r^2\right) = r_1^2 + r_2^2 \quad \text{(A-3)}$$

Law of cosine for $\triangle_{ABL}$:

$$r_1^2 + r_2^2 - 2r_1 r_2 \cos\alpha_{wp} = 4w_r^2 \quad \text{(A-4)}$$

$$(\text{A}-3),\ (\text{A}-4) : r_1 r_2 = \frac{r^2 - w_r^2}{\cos\alpha_{wp}} \quad \text{(A-5)}$$

Law of sine for $\triangle_{ACL}$: $\dfrac{\sin\alpha_1}{w_r} = \dfrac{\sin(180-\beta)}{r_1}$ (A-6)

Law of sine for $\triangle_{BCL}$: $\dfrac{\sin\alpha_2}{w_r} = \dfrac{\sin\beta}{r_2}$ (A-7)

$$(\text{A}-6) + (\text{A}-7) : 2\sin\beta = \frac{r_1\sin\alpha_1 + r_2\sin\alpha_2}{w_r} \quad \text{(A-8)}$$

Area of $\triangle$: $A_{ABL} = A_{ACL} + A_{BCL}$:

$$r_1 r_2 \sin\alpha_{wp} = r_1 r \sin\alpha_1 + r_2 r \sin\alpha_2 \quad \text{(A-9)}$$

$$(\text{A}-8),\ (\text{A}-9) : 2\sin\beta = \frac{r_1 r_2 \sin\alpha_{wp}}{r w_r} \quad \text{(A-10)}$$

$$(\text{A}-10),\ (\text{A}-5) : \sin\beta = \frac{r^2 - w_r^2}{2rw_r}\tan\alpha_{wp} \quad \text{(A-11)}$$

$\gamma = 90 - \beta$:

$$\gamma = \arccos\left(\frac{r^2 - w_r^2}{2rw_r}\tan\alpha_{wp}\right) \quad \text{(A-12)}$$

Using the following equations $\tan\alpha_{wp}$ can easily be determined (see Fig. 18):

$$\triangle_{EFL} : \tan\alpha = \frac{w_d}{f} \quad \text{(A-13)}$$

$$\triangle_{DFL} : \tan\left(\alpha + \alpha_{wp}\right) = \frac{w_p + w_d}{f} \quad \text{(A-14)}$$

Trigonometry :

$$\tan\left(\alpha + \alpha_{wp}\right) = \frac{\tan\alpha + \tan\alpha_{wp}}{1 - \tan\alpha\tan\alpha_{wp}} \quad \text{(A-15)}$$

$$(\text{A}-13),\ (\text{A}-14),\ (\text{A}-15) :$$

$$\tan\alpha_{wp} = \frac{f w_p}{f^2 + w_d\left(w_p + w_d\right)} \quad \text{(A-16)}$$

Since the landmarks are at a fixed height relative to the camera, this information can be used to determine the distance to the landmark (see Fig. 19).

$$(\text{A}-12) \text{ vertically} :$$

$$\cos\gamma_h = \arccos\left(\frac{r^2 - h_r^2}{2rh_r}\tan\alpha_{hp}\right)$$

$\triangle_{GHL}$: $\sin\gamma_h = \dfrac{z}{r}$

$\sin^2\gamma_h + \cos^2\gamma_h = 1$:

$$1 = z^2/r^2 + \frac{\left(r^2 - h_r^2\right)^2}{2rh_r}\tan^2\alpha_{hp}$$

$$r^2 = h_r^2\left[1 + 2\cot^2\alpha_{hp} \pm\sqrt{\left(1 + 2\cot^2\alpha_{hp}\right)^2 - \frac{h_r^2 + 4z^2\cot^2\alpha_{hp}}{h_r^2}}\right]$$

Finally, the three required parameters can be determined from the above equations:

$$R = \sqrt{r^2 - z^2}$$

$$\Gamma = \gamma - \alpha_{wc} + \delta_{cam} + 90$$

$$\theta = \delta_{cam} - \alpha_{wc}$$

References

1. Bobet P, Schmid C (1994) Obstacle Detection Analysis. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 796–799
2. Grimson WEL (1990) Object recognition by computer: the role of geometric constraints MIT Press, Cambridge
3. Lapreste J, Rives G Dhom M, Richetin M, (1989) Determination of the attitude of 3-D objects from a single perspective view IEEE PAMI 11:1265–1278
4. Kite DH, Magee M (1990) Determining the 3D position and orientation of a robot camera using 2D monocular vision Pattern Recognition. 23:819–831
5. Lazanas A, Latombe JC (1995) Motion planning with uncertainty: a landmark approach Artificial Intelligence 76:287–317
6. Gilg A Schmidt G (1994) Landmark-oriented visual navigation of a mobile robot IEEE Transactions on Industrial Electronics 41:392–397
7. Burl MC, Perona P (1996) Recognition of planar object classes In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, June, pp 223–230
8. Fukui I (1981) TV image processing to determine the position of a robot vehicle. Pattern Recognition 14:101–109
9. Courtney JW, Magee MJ, Aggarwal JK (1984) Robot guidance using computer vision Pattern Recognition 17:585–592

10. Aboaf EW, Drucker SM, Atkeson CG (1989) Task-level robot learning: juggling a tennis ball more accurately. IEEE International Conference on Robotics and Automation, Scottsdale, May, pp 1290–1295
11. De Geeter J (1996) A confidence set approach to localization. Technical report, GRASP Laboratory, University of Pennsylvania
12. Bajcsy R, Kamberova G, Mandelbaum R, Mintz M (1996) Robust fusion of position data. Workshop on Foundations of Information/Decision Fusion, Washington, August, pp 1–7
13. Kamberova G, Mintz M (1996) Robust minimax rules under zero-one loss for a restricted location parameter. Technical report 404, GRASP Laboratory, University of Pennsylvania
14. Mandelbaum R, Kamberova G, Mintz M (1996) Statistical decision theory for mobile robotics: theory and application. Technical report, GRASP Laboratory, University of Pennsylvania
15. Keppel G (1991) Design and analysis: a researcher's handbook. Prentice Hall, Englewood Cliffs
16. Geisser S, Greenhouse SW (1958) An extension of Box's results on the use of the F distribution in multivariate analysis. Annals of Mathematical Statistics 29:885–891
17. Mulligan IJ (1996) Empirical evaluation of information for robotic manipulation tasks. Thesis, University of British Columbia
18. Forstner W (1996) 10 Pros and Cons Against Performance Characterisation of Vision Algorithms. In Workshop on Performance Characteristics of Vision Algorithms, Cambridge, April

Péter L. Venetianer received the MS degree in computer science from the Technical University of Budapest in 1992 and the PhD degree in 1996. His doctoral work at the Computer and Automation Institute of the Hungarian Academy of Sciences focused on theoretical capabilities and applications of cellular neural/nonlinear networks. He is currently a Postdoctoral Fellow at the GRASP Laboratory at the University of Pennsylvania. His research interests include computer and machine vision, visual communication and coordination between multiple agents.

Edward Large studied mathematics and music performance as an undergraduate, receiving a diploma in mathematics from Southern Methodist University in 1982. He worked for several years in the United States and then in Japan before returning to graduate studies. His doctoral work at the Ohio State University concentrated on artificial intelligence, cognitive psychology, and the theory of computation. He received a PhD in Computer and Information Science in 1994. His dissertation centered on music perception and cognition and included both experimentation and modeling using nonlinear dynamical systems and neural networks. His research interests include the dynamics of human perception, action, and cognition, and the design of autonomous agents utilizing dynamical principles. He is currently a Postdoctoral Fellow at the Institute for Research in Cognitive Science and the GRASP Laboratory at the University of Pennsylvania.

Ruzena Bajcsy received a PhD in Electrical Engineering from the Slovak Technical University in Bratislava, Slovakia, in 1967 and a PhD in Computer Science from Stanford University in 1972. A faculty member of the University of Pennsylvania since 1972, Professor Bajcsy has covered the spectrum of problems involved in the field of computer vision. In 1983 she introduced the research paradigm of active perception, which connects perception with action. Recently she has been investigating cooperative behaviors between machines and humans. Professor Bajcsy's professional services to the community include: member of the NRC Computer Science and Telecommunication Board; advisory board member for the NSF Engineering and CISE Directorates; and member of the CRA Board of Directors. She is a Fellow of AAAI, ACM, IEEE, and the NAS Institute of Medicine.