Proceedings of the 1996 IEEE/SICE/RSJ International Conference on Multisensor Fusion and Integration for Intelligent Systems

Statistical Decision Theory for Mobile Robotics: Theory and Application *

G. Kamberova, R. Mandelbaum and M. Mintz
GRASP Laboratory
University of Pennsylvania, Philadelphia

Abstract

In this paper we pioneer a method which, given an input of mobile robot pose measurements by a sensor-based localization algorithm, produces a minimax risk fixed-size confidence set estimate for the pose of the agent. This work constitutes the first application to the mobile robotics domain of optimal fixed-size confidence-interval decision theory. The approach is evaluated in terms of theoretical capture probability and empirical capture frequency during actual experiments with the mobile agent. The method is compared to several other procedures including the Kalman Filter (minimum mean squared error estimate) and the Maximum Likelihood Estimator (MLE). The minimax approach is found to dominate all the other methods in performance. In particular, for the minimax approach, a very close agreement is achieved between theoretical *capture probability and* empirical *capture frequency. This allows performance to be accurately* predicted, *greatly facilitating the design of mobile robotic systems, and delineating the tasks that may be performed with a given system.*

1 Introduction

1.1 Overview

Inherent in the problem of mobile robotics is the requirement for the robot to estimate its pose (position and heading angle) as it changes over time. In this paper we develop efficient algorithms with guaranteed performance for set-valued pose estimation for mobile robots with uncertain dynamics which are subject to known geometric and algebraic constraints on both state and input variables. These algorithms are robust, i.e., able to contend with significant uncertainty in the noise distributions of the sensors. The novel aspects of our research are the determination of set-valued estimates of dynamic system behavior which apply: (i) where there is significant uncertainty in the underlying system dynamics and sensor noise behavior; (ii) where there are known geometric and algebraic constraints on both state and input variables; and (iii) where the estimation procedures are required to achieve probabilistic performance bounds. We adopt a set-valued basis for estimation, since this provides a method to obtain the desired probability statements, i.e., confidence bounds on the location of the object in pose space over a given time interval. Further, since the pose estimates are generally used in accomplishing some overall goal, e.g., multiagent motion coordination, it is important that the estimation modules provide performance guarantees which are amenable to these higher-level tasks. To obtain tight performance bounds, we select maximin probability statements as opposed to minimizing a functional such as mean squared error. Without the imposition of strong assumptions on the underlying sensor error distributions, as well as an accurate prior distribution for the uncertain parameters, moment-based methods such as mean squared error do *not* provide tight performance bounds. Tight performance bounds are critical components in the overall system operation and evaluation.

1.2 Salient features

1.2.1 Confidence set-based pose estimation

We use fixed-geometry confidence sets to obtain pose estimates of mobile agents with guaranteed performance bounds. The following assumptions and considerations apply:

(1) The uncertainty in the robot's initial pose is modeled geometrically. There is *no* need to assume an *a priori* probability distribution to model these uncertainties.

(2) The map features lie in given neighborhoods in the world coordinate system. Thus, we can account for survey errors. We do *not* require probabilistic models for these survey errors.

(3) The sensor error models are semiparametric and can be described by appropriate sets (envelopes) of sensor noise distributions. These uncertainty classes can be obtained empirically where nominal models do not suffice.

(4) The accumulated errors based on deadreckoning (or onboard inertial navigation systems) are described by set-valued dynamic models grounded in empirical data.

We have developed a mobile robot testbed system to test these concepts in an indoor environment. The agree-

*This research is funded in part by: ARPA Grants N00014-92-J-1647, DAAH04-93-G-0419; ARO Grants DAAL03-89-C-0031PRI, DAAL03-92-G0153; Gateway Grant 9109794; NASA Grants NGT-50729, NGT-70359; NIH Grant 3R01LM0521703S1; NSF Grants BCS92-16691, BCS92-21796, CISE/CDA-88-22719, CDA91-21973, CDA92-11136, CDA92-22732, GER93-55018, IRI92-10030, IRI92-09880, IRI93-03980, IRI93-07126, MSS91-57156-A 02; University Research Foundation Grant 370892; and The Whitaker Foundation

0-7803-3700-X/96 $5.00

ment between the *predicted capture probability* (confidence set coverage) based on our formal theory of robust fixed-geometry confidence sets and the *empirically determined system performance* is dramatic. Our ability to make these accurate predictions of system performance is of critical importance, since this makes it possible to meet *tight design specifications.*

1.2.2 Modular generic procedures

These estimation procedures are designed to be generic in that they can be applied to all autonomous mobile systems and combinations of sensor modalities.

Since the basic concepts for this approach combine geometric and probabilistic models which can be obtained empirically, these techniques do not require any specific hardware platform and can therefore be used in conjunction with all mobile bases and sensor suites.

1.3 Organization

In Section 2 we review basic decision theoretic concepts and summarize our theoretical contributions. Section 3 describes the experimental design for multidimensional confidence-set-based mobile robot pose estimation. Section 3.2 describes the method used for model-building and subsequent on-line confidence set estimation. The results for a sample trajectory are presented in Section 4. A discussion of these results follows in Section 5.

2 Theoretical background

2.1 Overview of the approach

Sensor errors can be modeled statistically, using both physical theory and empirical data. A sensor measurement often can be represented as a random variable $Z = \theta + V$ where V is random noise (error) with a distribution (CDF) F, and θ is the parameter of interest. In [3] we design optimal, in the minimax sense, decision rules for dealing with the uncertainties. These rules lead to confidence intervals for θ with guaranteed probabilities of capture. In the current paper we review these results.

This body of research, dealing with a continuous parameter, was initiated by Zeytinoglu and Mintz [8], [9], and later developed and extended by Kamberova [3]. McKendall [6] considers the corresponding problem for finite set of parameter values. In this section give some of the results from [3] relevant to the application presented in the current paper.

In developing the noise models, i.e., F, one recognizes that a single distribution is usually an inadequate description of sensor noise behavior. It is much more realistic and much safer to identify an envelope or class of distributions, one of whose members could reasonably represent the actual statistical behavior of a given sensor. The use of an *uncertainty class* in distribution space protects the system designer against the inevitable unpredictable changes which occur in sensor behavior. In [3] we look at three different types of uncertainty classes and design robust[1] minimax decision rules. The theory we address applies to a variety of uncertainty classes, which allows one to account for very general deviations from the nominal noise distribution. In this paper we introduce *tail-behavior uncertainty classes* which model well data with significant contamination in the tails of the distribution. This class is used in the experiment.

We develop the theory in the univariate case when θ is 1D. In order to utilize the method for practical mobile robot pose estimation, it is necessary to generalize from 1D confidence *interval* estimation to multivariate confidence *set* estimation. If we assume that the individual components of the multivariate parameter are approximately independent, we can apply the univariate analysis component-wise and then consider the confidence set which is the Cartesian product of the confidence intervals for the components [3]. In [4], Mandelbaum presents the first application of this theory to a mobile robotic setting, where the *minimax confidence set estimation (MCSE)* approach is used for mobile robot pose estimation. The empirical results accurately match those predicted by the theory, and dominate the performance of the commonly-used best linear procedures, namely the one-step Kalman filter. In the current paper some of the experimental results are highlighted and the method is compared with other existing techniques.

2.2 Decision-theoretic concepts

This section introduces decision-theoretic concepts and notation, and summarizes relevant decision-theoretic results. See [1] and [2] for a more detailed introduction to the subject of statistical-decision theory.

2.2.1 Location data model

Consider the *location data model* $Z = \theta + V$ where Z is a random variable representing the measurement, θ is the (unknown) true value of a parameter we wish to estimate, and V is random noise. We denote an observation (measurement) by z, $z \in \Re$. There is prior, non probabilistic information about θ given in terms of the *parameter space* Ω: $\theta \in \Omega = [-d, d]$. The random noise V has an absolutely continuous CDF F. The sampling distribution, the CDF of Z, is given by $F(z - \theta)$.

2.2.2 Decision rules, loss and risk functions

A *decision rule* $\delta : \Re \to \Omega$ is a function which given a measurement z produces an estimate $\delta(z)$.

A *loss function*, $L : \Re \times \Omega \to \Re$ is a function such that $L(\delta(z), \theta)$ corresponds to the "price" paid for the decision $\delta(z)$ when the true value of the parameter is θ. Common loss functions include *square error loss* $L(\delta(z), \theta) =$

[1] Robustness refers to the statistical effectiveness of the decision rule when the noise distribution is uncertain.

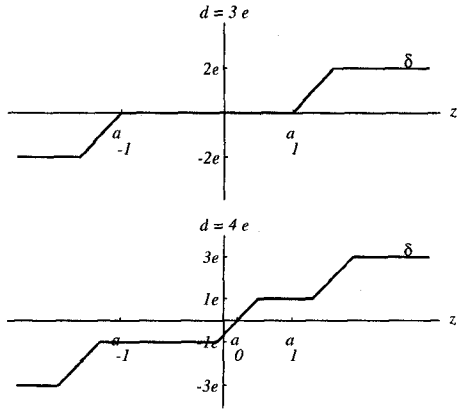

Figure 1: The structure of the minimax rules. Here, e denotes the error tolerance of the zero-one loss, L_e (e corresponds to ϵ elsewhere in this paper).

$(\delta(z) - \theta)^2$, and *zero-one loss with error tolerance* ϵ:

$$L_\epsilon(\delta(z), \theta) = \begin{cases} 0, & |\delta(z) - \theta| \le \epsilon; \\ 1, & |\delta(z) - \theta| > \epsilon; \end{cases}$$

If $\theta \in \Omega$, the *risk function* of a decision rule δ at θ is the expected loss when δ is used to estimate the parameter, i.e., $R(\delta, \theta, F) \stackrel{\text{def}}{=} \mathrm{E}_\theta^F L_\epsilon(\delta(Z), \theta) = \int_{-\infty}^{\infty} L_\epsilon(\delta(z), \theta)\, dF$.

2.2.3 Admissible and minimax rules

Decision rules are compared based on their *risk functions*. Given two rules, δ_1 and δ_2, we say that δ_1 is *better* that δ_2 if the risk of δ_1 is uniformly lower than that of δ_2, i.e. for any $\theta \in \Omega$, $R(\delta_1, \theta, F) \le R(\delta_2, \theta, F)$. A rule is *admissible* if there is no rule better than it.

A minimax rule δ^* is a rule which has minimal maximum risk, i.e., for any decision rule δ,

$$\sup_{\theta \in \Omega} R(\delta^*, \theta, F) \le \sup_{\theta \in \Omega} R(\delta, \theta, F).$$

Minimax rules are useful when there is no satisfactory prior probabilistic model for θ and when performance guarantees are needed to guard against worst possible errors. This is the case for mobile robot pose estimation presented here. The computation of minimax rules is a challenging problem. A major contribution of our research [3] is that we have proved the existence, delineated the structure, and presented a simple method for the computation of minimax rules for this given problem setting.

The minimax rules we obtain for the location-data model under zero-one loss have very specific structure. They are continuous, monotone nondecreasing, piecewise-linear functions with segments which have slope of 1 or 0. Such a rule is parameterized by a vector $\boldsymbol{a}$ which specifies the points where changes of slope occur. Figure 1 illustrates rules for the cases $d = 3\epsilon$ and $d = 4\epsilon$.

To find the minimax rule with the prescribed structure, we solve the risk equation, $R(\delta, \theta, F) = c$, for c and $\boldsymbol{a}$ where δ is parameterized by $\boldsymbol{a}$. In [3] we have proven that the solution to the risk equation exists and is unique. We have further derived the condition under which that solution gives an admissible minimax rule[2].

2.3 Minimax rules to confidence intervals

The attractiveness of the minimax rules under the location data model and zero-one loss stems from the connection between a minimax rule δ^* and a *fixed size confidence interval* for θ. Let $C^*(Z) \stackrel{\text{def}}{=} [\delta^*(Z) - \epsilon, \delta^*(Z) + e]$. For a location data model and zero-one loss,

$$R(\delta^*, \theta, F) = \Pr[\delta^*(Z) < \theta - \epsilon] + \Pr[\delta^*(Z) > \theta + \epsilon],$$

and $\Pr[\theta \in C^*(Z)] = 1 - R(\delta^*, \theta, F)$. Since δ^* is minimax, it follows that among all intervals of length 2ϵ, the interval $C^*(Z)$ can be interpreted as a confidence interval of size 2ϵ which has the highest confidence coefficient, defined as $\inf_{\theta,F} \Pr[\theta \in C^*(Z)]$. In other words, $C^*(Z)$ has the maximin probability of capturing θ within an interval of fixed length.

There is a trade-off between the value of the error-tolerance ϵ and the confidence coefficient of the obtained interval. The larger the value of ϵ, the higher the confidence coefficient, but the greater the uncertainty in the true value of θ. The value of ϵ is selected by the requirements of the application.

It should be also emphasized that minimax decision rules can be computed off-line, so that real-time estimation based on the confidence intervals is computationally efficient.

2.4 Uncertainty classes and robust rules

Often, it is safer and more realistic to assume that the noise is modeled not by a single distribution F but by a whole *uncertainty class* of distributions $\mathcal{F}$ to which the true distribution of V may belong. In this case, we seek decision rules which are robust against any distribution in the class $\mathcal{F}$.

A robust minimax rule, δ^*, for θ and $\mathcal{F}$, has lowest maximum risk, i.e., for any other rule δ and any $F \in \mathcal{F}$,

$$\sup_{\theta \in \Omega,\, F \in \mathcal{F}} R(\delta^*, \theta, F) \le \sup_{\theta \in \Omega,\, F \in \mathcal{F}} R(\delta, \theta, F).$$

The search for a *robust* minimax rule for $\mathcal{F}$ is based on the delineation of a minimax rule assuming noise with *a worst-case distribution for* $\mathcal{F}$. If we guard against errors under the worst-case model, we have a guaranteed performance for the whole class.

In [3] we study three different types of uncertainty classes and the worst-case distribution associated with each type. In this paper we focus on the *tail-behavior*

[2]The condition is that F be absolutely continuous with convex support, and that the density of F be such that $f(x + 2\epsilon)/f(x)$ is strictly monotone decreasing in x. Examples of such distributions include all distributions with monotone-likelihood ratio (such as Gaussian and Laplacian distributions), as well as non-symmetric and non-unimodal location-scale mixtures of normal distributions.

uncertainty class which is relevant to the practical application presented.

A strategy for deriving robust minimax rules for $\mathcal{F}$: First, delineate a worst-case distribution, F^*, for the class[3]. Second, compute the minimax rule assuming the noise has distribution F^*.

One of the major contributions of our work is the specification of uncertainty classes suitable for modeling practical applications, and the delineation of robust minimax rules for these classes [3]. In Section 2.5 we focus on the tail-behavior uncertainty class which has proven to be a suitable model for the task of mobile robot localization.

2.5 A tail-behavior uncertainty class

The *tail-behavior uncertainty class* $\mathcal{F}_{k_1,k_2,F_0}$ is depicted in Figure 2. $F \in \mathcal{F}_{k_1,k_2,F_0}$ coincides with F_0 within the

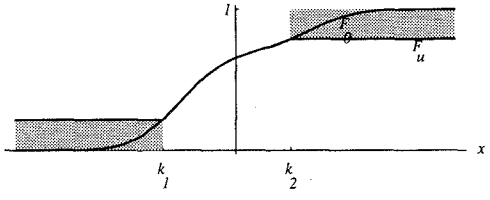

Figure 2: A tail-behavior uncertainty class

interval $[k_1, k_2]$, representing the range in sample space where measurements are reliable. Outside this interval, F can be any nondecreasing function whose graph lies in the shaded areas:

$$F[x] = \begin{cases} F_0[k_2] + (1 - F_0[k_2])H[x], & k_2 < x \\ F_0[x], & k_1 \leq x \leq k_2 \\ F_0[k_1]H[x], & x < k_1 \end{cases}$$

where H is an arbitrary CDF. In [3] we have shown that the worst-case distribution for $\mathcal{F}_{k_1,k_2,F_0}$ is

$$F_u(x) = \begin{cases} F_0[k_2], & k_2 \leq x; \\ F_0[x], & k_1 < x < k_2; \\ F_0[k_1], & x \leq k_1. \end{cases}$$

and and we have derived a robust minimax rule for the class. This is the rule which is used for robust pose estimation in Section 3. In that application, the nominal distribution for the tail-behavior uncertainty class is Gaussian and tail-behavior uncertainty arises from spurious sensor measurements. The system designer is spared the effort of explicitly modeling the probabilistic behavior of these measurements. Moreover, for practical applications, the empirical results presented in Section 4 confirm that the use of a tail-behavior uncertainty class and a robust rule is more suitable than traditional approaches which are based on a single distribution probability model.

[3]It is sufficient that F^* is absolutely continuous and has a density which has convex support.

2.6 Variations over time

In robotics environments, the size of the parameter set $2d$ and the observation noise distribution F may vary over time. If the behavior of d and F over time are known, it is possible to *predict* the probabilistic performance of fixed-size confidence interval estimation at any given time. This performance is measured in terms of the capture probability of the confidence interval. Knowledge of future probabilistic performance allows many design trade-offs to be addressed (see [4]):

(1) When should the agent stop moving and sensing the environment?

(2) Where should beacons be placed in the environment in a hybrid navigational system employing both artificially-inserted and naturally-occurring features? (Detection of a beacon resets d.)

(3) What level of performance can be expected given a certain geometric configuration of the environment and density of features, and hence what tasks may be attempted?

2.7 2D confidence-set-based estimation

In order to utilize the minimax decision rules for practical mobile robot pose estimation, it is necessary to generalize from 1D confidence *interval* estimation to multidimensional confidence *set* estimation. In [3], the results of [9] and [6] are generalized to the multivariate domain in the special case that the components of the parameter are approximately independently restricted, and the noise is independent among the components. As before, observations are assumed to be of the form $\boldsymbol{Z} = \boldsymbol{\theta} + \boldsymbol{V}$ where $\boldsymbol{Z}$ is now a multivariate measurement, $\boldsymbol{\theta}$ is the multivariate true parameter value, and $\boldsymbol{V}$ is multivariate additive observation noise. Once again, denote the (multivariate) distribution of $\boldsymbol{V}$ by F. Denote the multivariate parameter space by $\boldsymbol{\Omega}$. In the application of Section 3, the parameter vector we are estimating is *pose*. Hence $\boldsymbol{\Omega} \subset X \times Y \times \Phi$, where $X \times Y$ denotes the set of all possible *positions* of a mobile agent, and Φ denotes the set of all possible orientations.

The approach taken in the application of Section 3 is to assume that for each time step, F is the product of three independent 1D distributions, each aligned with a coordinate direction of some (unknown) coordinate frame $X' - Y' - \Phi$. This frame does not necessarily coincide with the global coordinate frame $X - Y - \Phi$, except in the Φ direction. We further assume that within the $X' - Y' - \Phi$ coordinate frame, the components of the parameter vector are independently restricted. The multivariate minimax confidence set is then computed as the Cartesian product of multiple confidence intervals, each aligned with one of the axes of $X' - Y' - \Phi$. The multivariate minimax capture probability is computed as the product of the three univariate capture probabilities along each component direction of $X' - Y' - \Phi$. Under these assumptions, all that is needed in order to generalize to the multivariate case is knowledge of the alignment of the $X' - Y' - \Phi$ coordinate

frame with respect to $X - Y - \Phi$. Section 3.2 describes how this alignment is derived empirically.

3 A confidence set approach to mobile robot localization

In this section we describe the apparatus and method used to implement the above theory for confidence-set-based pose estimation on a mobile robotic testbed. In addition, we describe our experimental setup, and the nature of the experiments we have performed.

3.1 Apparatus

The truth system: A critical element in the design and implementation of the fixed-size confidence set approach is a *truth system* for measuring the *true* pose of the robot at many points along a nominal trajectory. The truth system employed in these experiments is described in [4]. Note that this system is used solely for *model-building* and *evaluation* purposes; it is kept completely independent of all other systems on the testbed.

Logical sensor for pose measurement: We use an acoustic-based logical sensor to measure the pose of the agent at every time step. The measurement is based on a least-squares fit of numerous ultrasonic data collected by a ring of transducers over time. The fit is performed with respect to a given approximate geometric description of planar and corner features in the environment. See [5, 4] for details.

3.2 Method

The implementation of minimax confidence-set-based pose estimation on a mobile robotic testbed consists of three distinct phases: In the **model-building** phase, empirical time-varying models are built for the distribution F and the parameter set $\mathbf{\Omega}$. In the **rule-computation** phase, this location data model is used to compute the time-dependent minimax decision rules for pose estiamtion and their associated capture probabilities. These two phases are performed off-line. In the **estimation** phase, the decision rules are used to determine confidence sets for pose in real-time.

Model building: For a particular configuration of the environment, and for each nominal trajectory to be used, the following steps are used to derive an adequate model for the parameter set and noise distribution:

The mobile agent is made to traverse the nominal trajectory N times. For the jth run ($j = 1, 2, \ldots, N$), both the sequence of *true* poses $\{\boldsymbol{\theta}_1^j, \boldsymbol{\theta}_2^j, \ldots, \boldsymbol{\theta}_T^j\}$ and the sequence of *measured* poses $\{\boldsymbol{z}_1^j, \boldsymbol{z}_2^j, \ldots, \boldsymbol{z}_T^j\}$ are recorded, where T denotes the total number of time steps. For each time step t, the union of true poses $\bigcup_{j=1,2,\ldots,N} \boldsymbol{\theta}_t^j$ and of measured poses $\bigcup_{j=1,2,\ldots,N} \boldsymbol{z}_t^j$ is taken. In order to take into account a degree of asynchrony among the various runs, w consecutive sets of each sequence are grouped together and considered to belong to the same time *period*. The result is two sequences of sets $\bar{\boldsymbol{\theta}} = \{\bar{\boldsymbol{\theta}}_1, \bar{\boldsymbol{\theta}}_2, \ldots, \bar{\boldsymbol{\theta}}_{T/w}\}$ and $\bar{\boldsymbol{z}} = \{\bar{\boldsymbol{z}}_1, \bar{\boldsymbol{z}}_2, \ldots, \bar{\boldsymbol{z}}_{T/w}\}$.

Next, for each time period $\tau = 1, 2, \ldots, T/w$, extreme values are filtered from $\bar{\boldsymbol{z}}_\tau$ (they are considered to be outliers). A coordinate frame $X' - Y' - \Phi_\tau$ is found such that the scatter matrix of the remaining elements of $\bar{\boldsymbol{z}}_\tau$ is diagonal. An empirical distribution F_τ of observation error for time period τ is derived, based on the elements of $\bar{\boldsymbol{z}}_\tau$.

Similarly, for each time period $\tau = 1, 2, \ldots, T/w$, extreme values are filtered from $\bar{\boldsymbol{\theta}}_\tau$. The parameter set $\mathbf{\Omega}_\tau$ is taken to be the smallest rectangle aligned with the $X' - Y' - \Phi_\tau$ coordinate frame which contains all remaining elements of $\bar{\boldsymbol{\theta}}_\tau$.

Rule computation: For each time period τ, and for a given confidence interval half-width ϵ, the minimax decision rule δ_τ^* and corresponding risk R_τ are computed for each coordinate direction of $X' - Y' - \Phi'$. Each of the three coordinate directions is computed separately. The vector of rules $\boldsymbol{\delta}_\tau^*$ is used during estimation. The capture probability is computed as: $C_\tau = (1 - R_\tau^{X'})(1 - R_\tau^{Y'})(1 - R_\tau^{\Phi})$, where $R_\tau^{X'}$, $R_\tau^{Y'}$ and R_τ^{Φ} are the risks associated with the confidence interval in the coordinate directions X', Y' and Φ respectively.

Estimation: In this phase, a confidence set and corresponding capture probability C_τ are computed at each time step t of a traversal of the nominal trajectory. The placement of the confidence set depends on (a) the raw measurement $\boldsymbol{z}_t$ taken during the run, and (b) the pre-computed coordinate frame $X' - Y' - \Phi_\tau$, noise distribution F_τ, parameter set $\mathbf{\Omega}_\tau$, and vector of decision rules $\boldsymbol{\delta}_\tau^*$, where $\tau = \lceil t/w \rceil$. The center of the confidence set $\boldsymbol{c}_\tau$ is given by $\boldsymbol{c}_\tau = \boldsymbol{\omega}_\tau + \boldsymbol{\delta}_\tau^*(\Gamma_\tau(\boldsymbol{z}_t - \boldsymbol{m}_\tau))$ where $\boldsymbol{m}_\tau$ is the pre-computed mean of the noise distribution F_τ, $\boldsymbol{z}_t - \boldsymbol{m}_\tau$ denotes the observed pose in additive zero-mean noise, Γ_τ is the time-dependent rotation matrix which transforms a point from the $X - Y - \Phi_\tau$ into the $X' - Y' - \Phi_\tau$ coordinate system, and $\boldsymbol{\omega}_\tau$ is the center of the parameter set $\mathbf{\Omega}_\tau$. The three coordinate directions are treated independently; the confidence set is taken as the Cartesian product of the confidence intervals computed for each of the coordinate directions.

3.3 Evaluation criterion

The success of the confidence set approach outlined above is measured by whether the confidence set at time step t covers the true pose of the mobile agent at that time, as measured by the truth system of Section 3.1. For multiple runs, the frequency of capture at time step t may be calculated. Binning w frequencies for consecutive time steps yields the *empirical capture frequency (ECF)* ν_τ for time period $\tau = \lceil t/w \rceil$. If the modeled distribution F_τ accurately reflects the underlying real distribution for time period τ, then ν_τ will be approximately equal to the *theoretical capture probability (TCP)* C_τ for that time period.

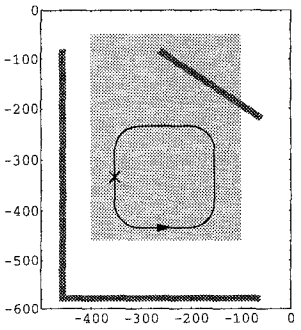

Figure 3: Layout of the environment. Thick black line-segments are planar reflecting surfaces. Light gray represents the region in which measurements of the *true* location of the agent can be made. Scale is in centimeters. Black line shows the nominal trajectory. The "X" marks the starting location. During each of 10 runs, the agent traversed the path twice in the direction indicated by the arrow.

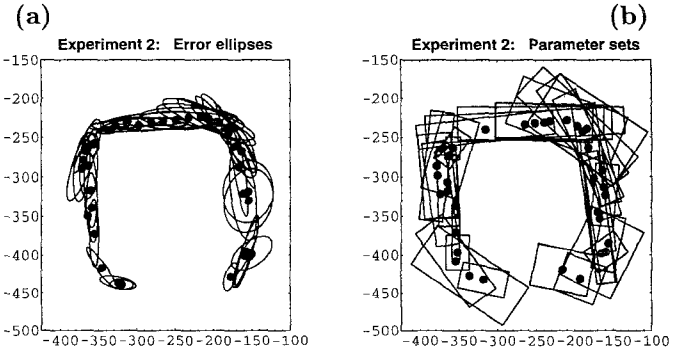

Figure 4: **(a)** Several second-order position error statistics S_t over time. The center of each ellipse marks the mean of the bivariate distribution, while the ellipse itself represents the standard deviation. The major and minor axes of all ellipses are magnified 5 times in this diagram. **(b)** Several position parameter sets $\mathbf{\Omega}_t$ over time. Each rectangle is aligned with the major axis of the corresponding error ellipse. The dimensions of all rectangles are magnified 5 times in this diagram. Scale is in centimeters.

3.4 Experimental setup

We have implemented the above approach and have examined its behavior for several different nominal trajectories using a fixed arrangement of features in the environment. In this paper we focus on one such trajectory, shown in Figure 3. The results for the other trajectories tested were qualitatively very similar [4].

For model-building, 10 traversals of the trajectory were made. The robot was run at a relatively slow speed of 50 millimeters per second since the truth system absorbs a large fraction of available computational resources (see Section 3.1). Time steps were spaced at approximately 250 milliseconds, with about 10 new ultrasonic measurements available at each time-step. Each traversal lasted for approximately 1060 time steps (265 seconds). Ultrasound-based localization [5] was treated as the measurement system. The steps outlined in Section 3.2 were then invoked in order to compute:

(1) For each time period τ, the mean $\boldsymbol{m}_\tau$ and scatter matrix S_τ of the observation error process $\boldsymbol{V}_\tau$, i.e. a second-order model. Note that the choice of a second-order model was motivated only by simplicity; the theory places no such restrictions on the model selected. A τ-dependent coordinate frame $X' - Y' - \Phi_\tau$ was selected such that S_τ was diagonal for each τ. Figure 4 gives intuition as to the order of magnitude of some of the covariance matrices associated with the noise process.

(2) A suitable parameter set $\mathbf{\Omega}_t = [\omega_\tau - \boldsymbol{d}_\tau, \omega_\tau + \boldsymbol{d}_\tau]$ for each time period with respect to the coordinate system $X' - Y' - \Phi_\tau$. In our case, data from all 10 runs were used to construct each parameter set. A time-binning factor of $w = 7$ was used, reflecting the uncertainty in synchronization of the various runs. Figure 4 gives intuition as to the order of magnitude of the parameter sets $\mathbf{\Omega}_\tau$.

The only two remaining components of the data model to be decided are (1) the distribution F_τ to model the noise process for each time period[4], and (2) the vector of half-dimensions $\boldsymbol{\epsilon}$ of the confidence set to be used. Once the data model is complete, we can compute the minimax decision rules for minimax confidence set estimation (MCSE) on future runs of the same nominal trajectory.

Where no ambiguity arises, rhe subscript τ will be suppressed in the sequel.

3.5 Noise models used in this experiment

The performance of the MCSE technique was analyzed and compared with respect to several choices of noise model. In addition, the performance of the technique was compared to that of a one-step Kalman Filter (KF) (i.e., the optimal[5] linear estimate based on a second-order model). In each case, the *theoretical* capture probability was compared to the *actual* frequency of capture. In all, the following approaches were compared:

(1) MCSE with Gaussian distribution $\mathcal{N}(\boldsymbol{m}, S)$.

(2) MCSE with distribution $\mathcal{N}(\boldsymbol{m}, 0.5625\,S)$ (i.e., the standard deviation used is 0.75 of the empirically measured standard deviation).

(3) MCSE with Gaussian distribution $\mathcal{N}(\boldsymbol{m}, 0.25\,S)$

(4) MCSE with double-sided exponential distribution $\mathcal{DS}\left(\boldsymbol{m}, \frac{1}{\sqrt{2}}S\right)$. Here the notation $\mathcal{DS}(\boldsymbol{m}, S)$ represents the product $\mathcal{DS}(m_1, S_{11}) \times \mathcal{DS}(m_2, S_{22}) \times \mathcal{DS}(m_3, S_{33})$ for S the diagonal matrix with diagonal elements S_{11}, S_{22} and S_{33}, and $\boldsymbol{m} = (m_1, m_2, m_3)^T$.

(5) *Robust* MCSE using the robust minimax rule for a tail-behavior uncertainty class in which the nominal distribution F_0 is the Gaussian $\mathcal{N}(\boldsymbol{m}, 0.25\,S)$, and $F_0(k_1) = \eta/2$ and $k_2 = -k_1$. See Section 2.5. The worst-case distribution for this class is a contaminated Gaussian distribution[6]. In this application we use $\eta = 0.1$.

[4] F_τ must have second-order statistics which match the empirically derived mean and scatter matrix for time period τ.

[5] with respect to expected squared error

[6] The *contamination* probability mass is assumed to constitute the fraction η of all the probability mass, and to be located sym-

(6) The Maximum Likelihood Estimator (MLE) $\delta(z) = z - m$. Theoretical capture probability was computed using a Gaussian distribution $\mathcal{N}(m, S)$. This approach corresponds to the MCSE technique when knowledge of the parameter sets Ω is ignored; i.e. the ratio d/ϵ is assumed to be infinite in all coordinate directions.

(7) One-step KF (optimal linear) estimation with jointly normal distributions $\mathcal{N}(m, S)$ and $\mathcal{N}(\mu, \Lambda)$, for the noise distribution and pose vector respectively. Under these assumptions, the one-step KF estimator is the best linear procedure and corresponds to the conditional expectation $E(\theta|z)$, where the conditional distribution $F_{\theta|z}$ has a Gaussian distribution $F_{\theta|z} \sim \mathcal{N}(v, \Upsilon)$, where $v = \mu + \Lambda(\Lambda + S)^{-1}(z - m)$, and $\Upsilon = \Lambda - \Lambda(\Lambda + S)^{-1}\Lambda$. In this document, we employ the *highest posterior density (HPD)* measure (see [1]) to compute the predicted capture probability for the KF model.

Note that *no* distribution needed to be assumed over the parameter space in the case of MCSE. Note also that in the case of MCSE, any plausible bivariate second-order distribution could have been selected to model observation noise. In the case of optimal linear (KF) estimation, on the other hand, an assumption of anything other than the Gaussian distribution would necessitate calculation of the conditional distribution $F_{\theta|z}$. See Section 5 for a discussion of the relative merits and constraints of the various approaches.

4 Results

Figure 5 illustrates a typical output of the MCSE position estimator for a single time step. A selected summary of results for *position* and *complete pose* estimation is presented in this paper. See [4] for a more detailed description of results.

For the remainder of this section, the merit of each approach for various distributions and values of ϵ will be summarized by the pair (ν, C) where C represents the theoretical capture probability (TCP) averaged over the ensemble of runs, and over all time periods. ν represents the mean *attained* frequency of capture (or *empirical capture frequency (ECF)*), again averaged over the ensemble and over time. An important measure of merit for each approach is not only the magnitudes of ν and C, but also the *closeness* of ν to C. A good match of the TCP and the ECF allows the TCP to be used as a *predictor* of performance for a given task. The ability to accurately forecast performance greatly facilitates system design.

4.1 Comparison of the seven approaches

In Figure 6 we compare the TCP and ECF for a single confidence set of size $2\epsilon = (10cm,\ 10cm,\ 2°)^T$ across the seven approaches outlined in Section 3.5 with respect to *position* estimation.

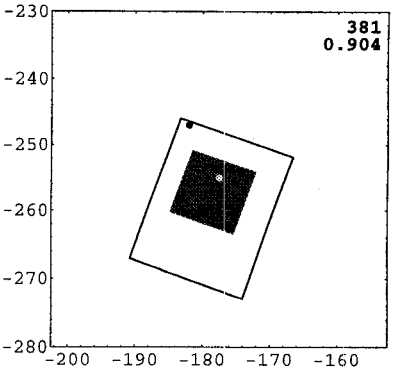

Figure 5: Typical output of the minimax confidence set position estimator for a single time step. The black rectangle defines the boundaries of the parameter set. It corresponds to the smallest rectangle which (1) is aligned with the coordinate system defined by the noise distribution, and (2) contains all true positions of the agent over multiple runs at approximately this time step. The light gray dot in the center of the diagram denotes the current true position of the agent, while the dark gray dot represents the current measurement of position. The grey square is the minimax confidence set for this time step. In this case, the confidence set has indeed captured the true position of the agent. The theoretical capture probability for this time step (number 381) is 0.904. Boundary units are in centimeters.

For each of the seven approaches, the closeness of the empirical to the theoretical capture frequency indicates how well the selected distribution models the underlying noise, and hence how reliably the TCP may be used during system design to *predict* performance for a given task. The MLE has by far the lowest theoretical capture probability. It also has the largest discrepancy between TCP and ECF, indicating a poor modeling of the underlying distribution. Its empirical capture frequency is comparable to some of the other approaches, however. For one-step Kalman Filter estimation, the TCP and ECF are close. However, the overall performance of the ECF is worst of all seven approaches. ECF performance of MSCE with distribution $\mathcal{N}(m, S)$ dominates that of one-step KF by approximately 15%. MCSE based on a double-sided exponential distribution performs approximately as well as MSCE based on a distribution $\mathcal{N}(m, S)$.

For MCSE based on Gaussian distributions, both TCP and ECF are progressively improved as the covariance matrix shrinks from S to $0.5625\,S$ and then to $0.25\,S$. However, the discrepancy between TCP and ECF also increases. The most favorable all-round performance is obtained for robust MCSE based on tail-behavior uncertainty class with a Gaussian nominal distribution: ECF and TCP are close, and both are high relative to the other approaches. Of the seven approaches analyzed, this data model most accurately reflects the true underlying behavior of the noise process. In [4] we examine several other trajectories. The results are similar.

metrically at $-\infty$ and ∞. The remaining mass is distributed as $\mathcal{N}(m, 0.25\,S)$. i.e. $F = \prod_{i=1}^{3}((1-\eta)\Psi_i + \eta/2)$, where $\Psi_i \sim \mathcal{N}(m_i, 0.25\,S_{ii})$ and S is the diagonal matrix with diagonal elements S_{11}, S_{22} and S_{33}.

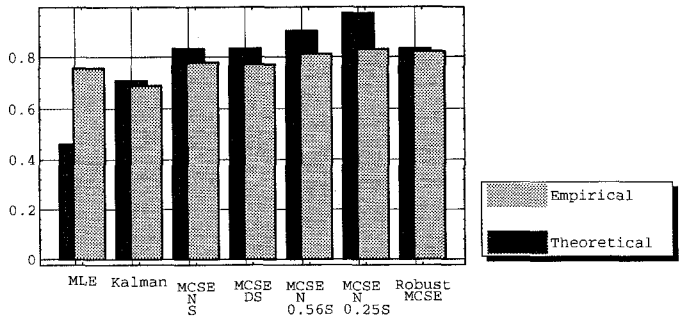

Figure 6: Comparison of seven approaches to *position* estimation for a confidence set of size $2\epsilon = (10cm,\ 10cm,\ 2\epsilon_\Phi)^T$. The seven labels denote the seven approaches listed in Section 3.4.

5 Discussion

For the MCSE approach, capture frequencies of approximately 0.6 are obtained for a confidence set of size 10cm × 10cm × 4°. We have tested MCSE and KF performance for different values of ϵ. The MCSE approach dominates the optimal linear estimator (one-step Kalman Filter) for both theoretical and empirical capture frequencies, and for all confidence set sizes tested. For a confidence set of size $2\epsilon = (10\text{cm}, 10\text{cm}, 4°)^T$, MCSE performance surpasses KF by as much as 50%. In addition to obtaining higher TCP and ECF values, the MCSE approach has the following attributes:

1. Non-probabilistic a priori geometric constraints on the unknown parameters (x, y and ϕ coordinates of pose in our case) may be incorporated. The geometric constraints are incorporated in the size of the parameter set $\mathbf{\Omega}$.

2. The MCSE technique is computationally tractable for a wide variety of noise distributions. Hence, the TCP and ECF can be brought into close alignment with a suitable model, enabling the TCP to be used as a *predictor* of performance for a given task. This ability to forecast performance *accurately* is an important factor in system design. The theory even handles semi-parametric *families* of distributions subject to extremely weak constraints. See [3].

3. A probabilistic dynamic model of the mobile robot is not required.

4. Once the model-building and rule-computation phases have been completed off-line, real-time estimation is computationally efficient.

In contrast, the optimal linear estimator (Kalman Filter) has the following difficulties:

1. The simplest and most common implementations of the optimal linear estimator assume linear stochastic models with Gaussian inputs. Such models do *not* constrain the robot's pose to a *compact set* in E^3. KF estimation using neutrally stable nonlinear models will also not suffice for incorporating geometric constraints. Moreover, nonlinear stochastic models with Gaussian inputs which *do* lead to probability distributions with compact support are often not adequate as dynamic models and are extremely difficult to employ in estimation procedures.

2. To employ *non-Gaussian* noise, and *non-linear* models in confidence set computations requires a complete analysis of the conditional probability distribution of the system state over time — in general, an exceptionally difficult computation.

3. In systems with *time-correlated* Gaussian observation noise, KF estimation requires the observation noise to be *whitened*, with **extra cost** in analysis and computation.

4. KF estimation is not as computationally efficient as MCSE.

6 Summary

We have delineated an approach to mobile robot pose estimation with **probabilistic performance guarantees** using fixed-geometry confidence sets. The technique is optimal against bounded location and a very wide range of observation noise distributions. The work described in this document constitutes the first application of the decision theory developed in [8], [9], [6], [7] and [3] to a mobile robotic setting.

The MCSE approach is very appropriate for the domain of mobile robot pose estimation. The empirical results confirm the utility of the theory and attest to the value of the MCSE procedure.

References

[1] J. O. Berger. *Statistical decision theory and Bayesian analysis.* Springer-Verlag, 1985.

[2] T. Ferguson. *Mathematical Statistics: a decision theoretic approach.* Academic Press, Inc., 1967.

[3] G. Kamberova. *Robust Location Estimation for MLR and Non-MLR Distributions.* PhD thesis, University of Pennsylvania, 1992.

[4] R. Mandelbaum. *Sensor Processing for Mobile Robot Localization, Exploration and Navigation.* PhD thesis, University of Pennsylvania, 1995.

[5] R. Mandelbaum and M. Mintz. Feature-based localization using fixed ultrasonic transducers. In *Proceedings of the IEEE International Conference on Intelligent Robots and Systems*, August 1995.

[6] R. McKendall. *Minimax estimation of a discrete location parameter for a continuous distribution.* PhD thesis, University of Pennsylvania, 1990.

[7] R. McKendall and M. Mintz. Data fusion techniques using robust statistics. In M. A. Abidi and R. C. Gonzalez, editors, *Data Fusion in Robotics and Machine Intelligence*, pages 211–244. Academic Press, 1992.

[8] M. Zeytinoglu and M. Mintz. Optimal fixed sized confidence procedures for a restricted parameter space. *The Annals of Statistics*, 12:945–957, September 1984.

[9] M. Zeytinoglu and M. Mintz. Robust optimal fixed sized confidence procedures for a restricted parameter space. *The Annals of Statistics*, 16:1241–1253, September 1988.