IEICE TRANS. INF. & SYST., VOL. E81–D, NO. 0 1998

PAPER

Optimal Robot Self-Localization and Accuracy Bounds *

Kenichi KANATANI† *and* **Naoya OHTA**†, *Members*

SUMMARY We discuss optimal estimation of the current location of a mobile robot by matching an image of the scene taken by the robot with the model of the environment. We first present a theoretical accuracy bound and then give a method that attains that bound, which can be viewed as describing the probability distribution of the current location. Using real images, we demonstrate that our method is superior to the naive least-squares method. We also confirm the theoretical predictions of our theory by applying the bootstrap procedure.

key words: mobile robot, image sensing, self-localization, statistical estimation, theoretical accuracy bound, bootstrap

1. Introduction

For a mobile robot to navigate autonomously, it must have a geometric model of the environment; it may be given as data or constructed by the robot itself from vision and sensor data. Here, we consider the case in which the robot already has a three-dimensional map of the environment and study the problem of identifying its current location in the world model. In theory, the current location can be computed by tracing the history of motion from a known initial position, e.g., integrating the rotation of the wheels or incrementally correcting the position by estimating robot motion from images [8]. However, the accuracy of the computed location quickly deteriorates as errors (due to slippage of the wheels, vibration of the camera, etc.) are accumulated in the course of integration. At some point, therefore, we need to estimate the current location by some direct means.

A typical method for self-localization is computing the current position of the camera by matching feature points detected in the images with their corresponding positions in the world model. The most direct method is stereo vision, by which the 3-D locations of the feature points relative to the cameras can be computed [1]. This fails, however, if the feature points are located very far away as compared with the baseline of the stereo system. In an outdoor environment, feature points easily discernible from a wide range of positions are usually those located very far away (e.g., towers and mountain tops). Hence, we need a method for computing the current position by matching a single image with the world model.

The problem of matching image and model features is inseparable from the problem of computing the 3-D position; we first hypothesize a matching based on known clues (e.g., brightness, color, shape, etc.) and then validate the resulting 3-D position (e.g., by comparing it with that obtained by integrating the history of motion, examining the image if features that should be observed from that position actually exist, etc.) [10], [11]. Hence, computing the 3-D position for given matching between image and model features is crucial whether the matching is correct or not. However, the actual matching procedure depends on the situation to be applied, because what kind of feature to detect depends on the nature of the scene.

Computing the 3-D relationship between image and model features has been studied by many researchers in the past in the form known as "PnP", in which the goal is to compute the 3-D positions of the feature points relative to the camera, given a 3-D configuration of the feature points relative to each other. Here, we are interested in computing the absolute position of the camera, given absolute 3-D positions of feature points.

If the robot motion is constrained to be on a horizontal surface (e.g., the ground or a floor), a simple method based on elementary geometry of circles is well known for this purpose [7]. It can also be applied to three-dimensional motion by replacing circles by spheres [9]. But this technique uses only pairwise relative orientations of the lines of sight defined by the feature points; their absolute positions in the image are not used. Using minimal information has the advantage that it can be adapted to mismatch removal: we pick out multiple minimal sets of data and choose the solution supported by majority voting [4]. For assumed matching, however, it is obviously better to fuse all available information in an optimal manner. Such a method also exists [2], but so far the main concern has been *methods* for estimation; little attention has been paid to *theoretical optimality* and *reliability of the solution*.

The aim of this paper is *not* to propose yet another new solution technique. Rather, we focus on *statistical*

Manuscript received February 2, 1998.
Manuscript revised June *, 1998.

†The authors are with the Department of Computer Science, Gunma University, Kiryu-shi, 376-8515 Japan.

*This work was in part supported by the Ministry of Education, Science, Sports and Culture, Japan under a Grant in Aid for Scientific Research C(2) (No. 09680352).

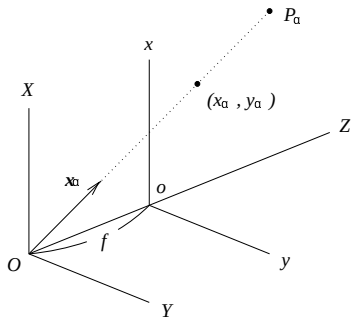


Fig. 1 Camera imaging geometry.

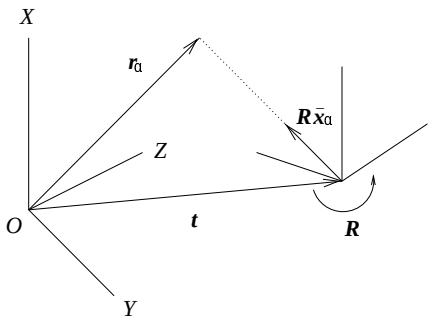


Fig. 2 The camera coordinate system and the world coordinate system.

aspects. We first introduce a model of noise and view the problem as statistical estimation. Then, we present a *theoretical accuracy bound* that can be evaluated independently of particular solution techniques involved. Next, we give a computational scheme that attains that bound; such a method alone can be called "optimal".

Since the solution attains the accuracy bound, we can view it as quantitatively describing the "probability distribution" of the current location of the robot. We show that we can compute this distribution without any knowledge about the magnitude of image noise. This computation helps validate the hypothesized matching: if the evaluated distribution spreads out widely, the hypothesis is very questionable. We confirm the theoretical predictions of our theory by using real images and applying the bootstrap procedure [3].

2. Statistical Self-Localization

We regard the camera imaging geometry as perspective projection and define an XYZ camera coordinate system in such a way that its origin is at the center of projection and its optical axis is along the Z-axis (Fig. 1). Letting f be the focal length, we identify the plane $Z = f$ with the image plane, on which we define an xy image coordinate system in such a way that the origin is on the optical axis and the x- and y-axes are parallel to the X- and Y-axes, respectively.

We regard observed image coordinates (x_α, y_α) (in pixels) as perturbed from their true values $(\bar{x}_\alpha, \bar{y}_\alpha)$ by noise and write

$$x_\alpha = \bar{x}_\alpha + \Delta x_\alpha, \qquad y_\alpha = \bar{y}_\alpha + \Delta y_\alpha. \tag{1}$$

We regard Δx_α and Δy_α as (generally correlated) Gaussian random variables of mean 0, independent for each α.

Suppose the camera coordinate system is in a position defined by translating the world coordinate system by $\boldsymbol{t}$ and rotating it by $\boldsymbol{R}$ with respect to the world coordinate system (Fig. 2). We call $\{\boldsymbol{t}, \boldsymbol{R}\}$ the *motion parameters*. Our goal is formally stated as follows:

Problem 1: Given image coordinates (x_α, y_α), $\alpha = 1, ..., N$, of feature points whose *3-D* positions $\boldsymbol{r}_\alpha$, $\alpha = 1, ..., N$, with respect to the world coordinate system are known, optimally compute the motion parameters $\{\boldsymbol{t}, \boldsymbol{R}\}$ and their probability distribution.

We represent a point with image coordinates (x, y) by vector. $\boldsymbol{x} = (x/f, y/f, 1)^\top$. This vector indicates the line of sight starting from the camera coordinate origin and passing through the corresponding point in the scene (Fig. 1). Let $\boldsymbol{x}_\alpha$ and $\bar{\boldsymbol{x}}_\alpha$ be the αth observed point and its true position, respectively. The error $\Delta\boldsymbol{x}_\alpha = \boldsymbol{x}_\alpha - \bar{\boldsymbol{x}}_\alpha$ is a three-dimensional vector. We define its covariance matrix by

$$V[\boldsymbol{x}_\alpha] = E[\Delta\boldsymbol{x}_\alpha \Delta\boldsymbol{x}_\alpha^\top], \tag{2}$$

where $E[\cdot]$ denotes expectation and the superscript $\top$ denotes transpose. Since the Z component of $\Delta\boldsymbol{x}_\alpha$ is identically 0, the covariance matrix $V[\boldsymbol{x}_\alpha]$ is singular; its third row and third column consist of 0s.

The covariance matrix $V[\boldsymbol{x}_\alpha]$ measures the uncertainty of detecting the feature point $\boldsymbol{x}_\alpha$, but in practice it is usually very difficult to predict it precisely. However, it is often possible to predict the relative likelihood of noise. Here, we assume that the covariance matrix is known only *up to scale* and write

$$V[\boldsymbol{x}_\alpha] = \epsilon^2 V_0[\boldsymbol{x}_\alpha]. \tag{3}$$

We assume that $V_0[\boldsymbol{x}_\alpha]$ is known but the constant ϵ is unknown; we call ϵ the *noise level*, and $V_0[\boldsymbol{x}_\alpha]$ (generally different from point to point) the *normalized covariance matrix* [6]. For example, if Δx_α and Δy_α are subject to an isotropic and identical Gaussian distribution of mean 0 and variance σ^2, we have

$$\epsilon = \frac{\sigma}{f}, \qquad V_0[\boldsymbol{x}_\alpha] = \mathrm{diag}(1, 1, 0), \tag{4}$$

where $\mathrm{diag}(\lambda_1, \lambda_2, \lambda_3)$ denotes the diagonal matrix with diagonal elements λ_1, λ_2, and λ_3 in that order.

The vector $\bar{\boldsymbol{x}}_\alpha$ is defined with respect to the camera coordinate system. If it is described with respect to the world coordinate system, it becomes $\boldsymbol{R}\bar{\boldsymbol{x}}_\alpha$ (Fig. 2). Hence, letting Z_α be the depth of the αth feature point in the scene from the camera coordinate origin, we obtain the following relationship:

$$\boldsymbol{r}_\alpha = \boldsymbol{t} + Z_\alpha \boldsymbol{R}\bar{\boldsymbol{x}}_\alpha. \tag{5}$$

Such a depth Z_α exists if and only if vector $\boldsymbol{r}_\alpha - \boldsymbol{t}$ is parallel to vector $\boldsymbol{R}\bar{\boldsymbol{x}}_\alpha$. Hence, Problem 1 reduces to the following statistical estimation:

Problem 2: Given $\{\boldsymbol{r}_\alpha\}$, estimate the motion parameters $\{\boldsymbol{t}, \boldsymbol{R}\}$ that satisfy

$$(\boldsymbol{t} - \boldsymbol{r}_\alpha) \times \boldsymbol{R}\bar{\boldsymbol{x}}_\alpha = \boldsymbol{0}, \qquad \alpha = 1, ..., N, \tag{6}$$

from the noisy data $\{\boldsymbol{x}_\alpha\}$. At the same time, compute the probability distribution of the estimated motion parameters $\{\boldsymbol{t}, \boldsymbol{R}\}$.

3. Theoretical Accuracy Bound

Let $\{\hat{\boldsymbol{t}}, \hat{\boldsymbol{R}}\}$ be an estimator of the true motion parameters $\{\bar{\boldsymbol{t}}, \bar{\boldsymbol{R}}\}$. The deviation of translation can be measured by the "difference" $\Delta\boldsymbol{t} = \hat{\boldsymbol{t}} - \bar{\boldsymbol{t}}$ of the estimator $\hat{\boldsymbol{t}}$ from its true value $\bar{\boldsymbol{t}}$. The deviation of rotation can be measured by the "quotient" $\hat{\boldsymbol{R}}\bar{\boldsymbol{R}}^\top$, i.e., the rotation of $\hat{\boldsymbol{R}}$ relative to $\bar{\boldsymbol{R}}$. Let $\boldsymbol{l}$ (unit vector) and $\Delta\Omega$ be, respectively, the axis and angle of the relative rotation $\hat{\boldsymbol{R}}\bar{\boldsymbol{R}}^\top$. We let $\Delta\boldsymbol{\Omega} = \Delta\Omega\boldsymbol{l}$ and define the covariance matrices of the estimator $\{\hat{\boldsymbol{t}}, \hat{\boldsymbol{R}}\}$ as follows:

$$\begin{aligned} V[\hat{\boldsymbol{t}}] &= E[\Delta\boldsymbol{t}\Delta\boldsymbol{t}^\top], \quad V[\hat{\boldsymbol{t}}, \hat{\boldsymbol{R}}] = E[\Delta\boldsymbol{t}\Delta\boldsymbol{\Omega}^\top], \\ V[\hat{\boldsymbol{R}}] &= E[\Delta\boldsymbol{\Omega}\Delta\boldsymbol{\Omega}^\top], \quad V[\hat{\boldsymbol{R}}, \hat{\boldsymbol{t}}] = E[\Delta\boldsymbol{\Omega}\Delta\boldsymbol{t}^\top]. \end{aligned} \tag{7}$$

Applying the theory of Kanatani [6], we can obtain the following lower bound, which Kanatani called the *Cramer-Rao lower bound* in analogy with the corresponding bound in traditional statistical estimation:

$$\begin{pmatrix} V[\hat{\boldsymbol{t}}] & V[\hat{\boldsymbol{t}}, \hat{\boldsymbol{R}}] \\ V[\hat{\boldsymbol{R}}, \hat{\boldsymbol{t}}] & V[\hat{\boldsymbol{R}}] \end{pmatrix} \succ \epsilon^2 \times \begin{pmatrix} \sum_{\alpha=1}^N \bar{\boldsymbol{A}}_\alpha^\top \bar{\boldsymbol{W}}_\alpha \bar{\boldsymbol{A}}_\alpha & \sum_{\alpha=1}^N \bar{\boldsymbol{A}}_\alpha^\top \bar{\boldsymbol{W}}_\alpha \bar{\boldsymbol{B}}_\alpha \\ \sum_{\alpha=1}^N \bar{\boldsymbol{B}}_\alpha^\top \bar{\boldsymbol{W}}_\alpha \bar{\boldsymbol{A}}_\alpha & \sum_{\alpha=1}^N \bar{\boldsymbol{B}}_\alpha^\top \bar{\boldsymbol{W}}_\alpha \bar{\boldsymbol{B}}_\alpha \end{pmatrix}^{-1}. \tag{8}$$

Here, $\boldsymbol{U} \succ \boldsymbol{V}$ means that $\boldsymbol{U} - \boldsymbol{V}$ is a positive semidefinite symmetric matrix, and matrices $\bar{\boldsymbol{A}}_\alpha$, $\bar{\boldsymbol{B}}_\alpha$, and $\bar{\boldsymbol{W}}_\alpha$ are defined as follows ($\boldsymbol{I}$ is the unit matrix):

$$\begin{aligned} \bar{\boldsymbol{A}}_\alpha &= -(\bar{\boldsymbol{R}}\bar{\boldsymbol{x}}_\alpha) \times \boldsymbol{I}, \\ \bar{\boldsymbol{B}}_\alpha &= (\bar{\boldsymbol{t}}_\alpha - \boldsymbol{r}_\alpha, \bar{\boldsymbol{R}}_\alpha\bar{\boldsymbol{x}}_\alpha)\boldsymbol{I} - \bar{\boldsymbol{R}}\bar{\boldsymbol{x}}_\alpha(\bar{\boldsymbol{t}} - \boldsymbol{r}_\alpha)^\top, \end{aligned} \tag{9}$$

$$\bar{\boldsymbol{W}}_\alpha = \Big((\bar{\boldsymbol{t}} - \boldsymbol{r}_\alpha) \times \bar{\boldsymbol{R}}V_0[\boldsymbol{x}_\alpha]\bar{\boldsymbol{R}}^\top \times (\bar{\boldsymbol{t}} - \boldsymbol{r}_\alpha)\Big)^-. \tag{10}$$

Throughout this paper, the inner product of vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ is denoted by $(\boldsymbol{u}, \boldsymbol{v})$. The product $\boldsymbol{v} \times \boldsymbol{U}$ of a vector $\boldsymbol{v}$ and a matrix $\boldsymbol{U}$ is the matrix whose columns are the vector products of $\boldsymbol{v}$ and the columns of $\boldsymbol{U}$. The product $\boldsymbol{U} \times \boldsymbol{v}$ of a matrix $\boldsymbol{U}$ and a vector $\boldsymbol{v}$ is the matrix whose rows are the vector products of the rows of $\boldsymbol{U}$ and vector $\boldsymbol{v}$. The operation $(\cdot)^-$ designates the (*Moore-Penrose*) *generalized inverse*.

4. Optimal Estimation

Applying the general theory of Kanatani [6], we can obtain a computational scheme for solving Problem 2 in such a way that the resulting solution attains the accuracy bound (8) in the first order (i.e., ignoring terms of $O(\epsilon^4)$): we minimize the sum of squared *Mahalanobis distances*

$$J = \sum_{\alpha=1}^N (\bar{\boldsymbol{x}}_\alpha - \boldsymbol{x}_\alpha, V_0[\boldsymbol{x}_\alpha]^-(\bar{\boldsymbol{x}}_\alpha - \boldsymbol{x}_\alpha)) \tag{11}$$

with respect to $\{\bar{\boldsymbol{x}}_\alpha\}$ subject to the constraint (6). Introducing Lagrange multipliers to the constraint (6) and eliminating $\bar{\boldsymbol{x}}_\alpha$, we can rewrite Eq. (11) as follows:

$$J = \sum_{\alpha=1}^N ((\boldsymbol{t} - \boldsymbol{r}_\alpha) \times \boldsymbol{R}\boldsymbol{x}_\alpha, \boldsymbol{W}_\alpha \Big((\boldsymbol{t} - \boldsymbol{r}_\alpha) \times \boldsymbol{R}\boldsymbol{x}_\alpha\Big)). \tag{12}$$

The matrix $\boldsymbol{W}_\alpha$ is defined by

$$\boldsymbol{W}_\alpha = \Big((\boldsymbol{t} - \boldsymbol{r}_\alpha) \times \boldsymbol{R}V_0[\boldsymbol{x}_\alpha]\boldsymbol{R}^\top \times (\boldsymbol{t} - \boldsymbol{r}_\alpha)\Big)_2^-, \tag{13}$$

where the operation $(\cdot)_r^-$ designates the *rank-constrained* (*Moore-Penrose*) *generalized inverse* computed by transforming it into the canonical form, replacing its eigenvalues except the r largest ones by 0, and computing the (Moore-Penrose) generalized inverse; this operation is necessary for preventing numerical instability [6].

Let $\hat{J}$ be the *residual*, i.e., the minimum of J. Since $\hat{J}/\epsilon^2$ is subject to a χ^2 distribution with $2N-6$ degrees of freedom in the first order [6], we obtain an unbiased estimator of the squared noise level ϵ^2 in the following form:

$$\hat{\epsilon}^2 = \frac{\hat{J}}{2N - 6}. \tag{14}$$

In other words, the unknown noise level ϵ can be estimated *a posteriori* by measuring the discrepancy between the observed image and the model image viewed from the optimally computed position. Hence, all sources of errors including camera calibration errors, 3-D modeling errors, and mismatches of feature points are regarded as "noise".

Because the solution $\{\hat{\boldsymbol{t}}, \hat{\boldsymbol{R}}\}$ of the minimization (12) attains the accuracy bound (8) in the first order, we can evaluate their covariance matrices by substituting the data $\{\boldsymbol{x}_\alpha\}$, the computed solution $\{\hat{\boldsymbol{t}}, \hat{\boldsymbol{R}}\}$, and the estimate $\hat{\epsilon}^2$ computed by eq.(14) for their true values $\{\bar{\boldsymbol{x}}_\alpha\}$, $\{\bar{\boldsymbol{t}}, \bar{\boldsymbol{R}}\}$, and ϵ^2, respectively, in Eqs. (8). Using the covariance matrix $V[\hat{\boldsymbol{t}}]$ in Eqs. (8), we can estimate the probability distribution of the current location in the following form:

$$p(\boldsymbol{r}) = \frac{1}{(2\pi|V[\hat{\boldsymbol{t}}]|)^{3/2}} e^{-(\boldsymbol{r}-\hat{\boldsymbol{t}}, V[\hat{\boldsymbol{t}}]^{-1}(\boldsymbol{r}-\hat{\boldsymbol{t}}))/2}. \quad (15)$$

We conduct the minimization (12) by modified Newton iterations. If rotation $\boldsymbol{R}$ is perturbed by a small rotation represented by the vector $\Delta\boldsymbol{\Omega}$, the perturbed rotation has the expression

$$\boldsymbol{R} + \Delta\boldsymbol{\Omega} \times \boldsymbol{R} + \frac{1}{2}\Delta\boldsymbol{\Omega}\Delta\boldsymbol{\Omega}^\top\boldsymbol{R} - \frac{1}{2}\|\Delta\boldsymbol{\Omega}\|^2\boldsymbol{R} + O(\Delta\boldsymbol{\Omega})^3, \quad (16)$$

where $\|\boldsymbol{u}\|$ denotes the norm of a vector $\boldsymbol{u}$ and $O(\boldsymbol{u}, \boldsymbol{v}, ...)^k$ designates terms of order k or higher in the elements of vectors $\boldsymbol{u}$, $\boldsymbol{v}$, ... Substituting Eq. (16) and $\boldsymbol{t} + \Delta\boldsymbol{t}$ for $\boldsymbol{R}$ and $\boldsymbol{t}$, respectively, in Eq. (12) and expanding it with respect to $\Delta\boldsymbol{t}$ and $\boldsymbol{\Omega}$, we obtain an expression in the form

$$J + (\nabla_t J, \Delta\boldsymbol{t}) + (\nabla_R J, \Delta\boldsymbol{\Omega}) + \frac{1}{2}(\Delta\boldsymbol{t}, \nabla^2_{tt} J, \Delta\boldsymbol{t}) + (\Delta\boldsymbol{t}, \nabla^2_{tR} J, \Delta\boldsymbol{\Omega}) + \frac{1}{2}(\Delta\boldsymbol{\Omega}, \nabla^2_{RR} J, \Delta\boldsymbol{\Omega}) + O(\Delta\boldsymbol{t}, \Delta\boldsymbol{\Omega})^3. \quad (17)$$

Differentiating this with respect to $\Delta\boldsymbol{t}$ and $\Delta\boldsymbol{\Omega}$, letting the resulting expressions equal zero, and ignoring terms of $O(\Delta\boldsymbol{t}, \Delta\boldsymbol{\Omega})^3$, we obtain simultaneous linear equations in the form

$$\begin{pmatrix} \nabla^2_{tt} J & \nabla^2_{tR} J \\ (\nabla^2_{tR} J)^\top & \nabla^2_{RR} J \end{pmatrix} \begin{pmatrix} \Delta\boldsymbol{t} \\ \Delta\boldsymbol{\Omega} \end{pmatrix} = -\begin{pmatrix} \nabla_t J \\ \nabla_R J \end{pmatrix}. \quad (18)$$

Starting from an initial guess $\{\boldsymbol{t}, \boldsymbol{R}\}$, we solve Eq. (18) for the increments $\{\Delta\boldsymbol{t}, \Delta\boldsymbol{\Omega}\}$ and update the solution in the form $\boldsymbol{t} \leftarrow \boldsymbol{t} + \Delta\boldsymbol{t}$ and $\boldsymbol{R} \leftarrow \mathcal{R}(\Delta\boldsymbol{\Omega})\boldsymbol{R}$, where $\mathcal{R}(\Delta\boldsymbol{\Omega})$ designates the rotation matrix of angle $\|\Delta\boldsymbol{\Omega}\|$ around axis $\Delta\boldsymbol{\Omega}$. We iterate this until $\|\Delta\boldsymbol{t}\| < \epsilon_t$ and $\|\Delta\boldsymbol{\Omega}\| < \epsilon_R$ for specified thresholds ϵ_t and ϵ_R.

We compute the initial guess $\{\boldsymbol{t}, \boldsymbol{R}\}$ by a *structure-from-motion* algorithm. First, we hypothetically place a reference camera coordinate system in a known position in the world model and compute the image coordinates of the feature points viewed from that position. From the correspondences of image coordinates between this reference image and the actually observed image, we then reconstruct the 3-D motion of the camera and the 3-D positions of the feature points up to scale; since we know the absolute positions of the feature points, we can adjust the scale a posteriori. Here, we adopt the statistically optimal algorithm of Kanatani [5] using a technique called *renormalization*.

5. Examples

Fig. 3 is a real image of a toy house. We manually input the feature points marked by white dots and used the noise model of Eqs. (4). The computation converged after five iterations for thresholds $\epsilon_t = 0.01$cm (the height of the house is 8cm) and $\epsilon_R = 0.01°$. Fig. 4 displays the house and the estimated camera coordinate axes viewed from an angle.

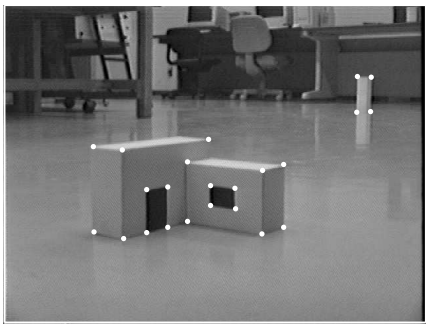


Fig. 3 A real image of a toy house.

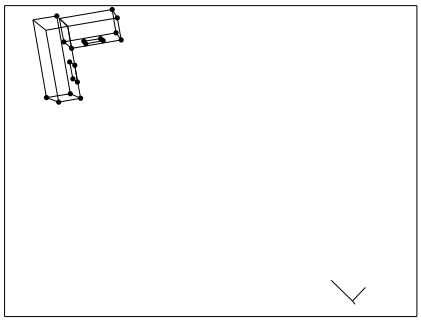


Fig. 4 Estimated current location.

We evaluated the reliability of the computed solution $\{\hat{\boldsymbol{t}}, \hat{\boldsymbol{R}}\}$ by applying a procedure called *bootstrap* [3]. We first optimally correct the observed positions $\{\boldsymbol{x}_\alpha\}$ into the positions $\{\hat{\boldsymbol{x}}_\alpha\}$ that exactly satisfy the constraint (6). This optimal correction is done as follows:

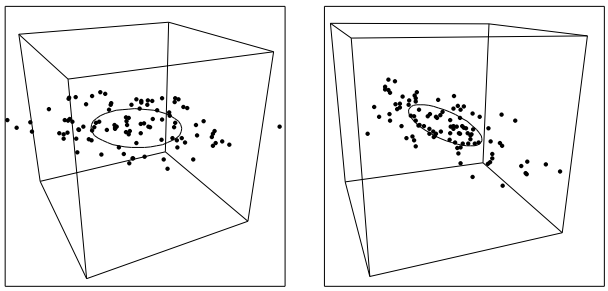
$$\hat{\boldsymbol{x}}_\alpha = \boldsymbol{x}_\alpha - V_0[\boldsymbol{x}_\alpha]\hat{\boldsymbol{R}}^\top \Big((\hat{\boldsymbol{t}} - \boldsymbol{r}_\alpha) \times \hat{\boldsymbol{W}}_\alpha \times (\hat{\boldsymbol{t}} - \boldsymbol{r}_\alpha) \Big) \hat{\boldsymbol{R}}\boldsymbol{x}_\alpha. \quad (19)$$

Here, $\hat{\boldsymbol{W}}_\alpha$ is the value of $\boldsymbol{W}_\alpha$ obtained by replacing $\{\boldsymbol{t}, \boldsymbol{R}\}$ by $\{\hat{\boldsymbol{t}}, \hat{\boldsymbol{R}}\}$ in Eq. (13).

Estimating the noise variance by Eq. (14), we generate random Gaussian noise that has the estimated variance and add it to the corrected positions independently. Then, we compute the motion parameters $\{\boldsymbol{t}^*, \boldsymbol{R}^*\}$ and the angle $\Delta\Omega^*$ and axis $\boldsymbol{l}^*$ of the relative rotation $\hat{\boldsymbol{R}}^*\hat{\boldsymbol{R}}^\top$.

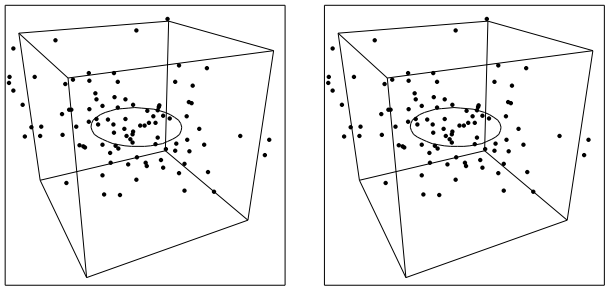
Fig. 5 shows three-dimensional plots of the error vectors $\Delta\boldsymbol{t}^* = \boldsymbol{t}^* - \hat{\boldsymbol{t}}$ and $\Delta\boldsymbol{\Omega}^* = \Delta\Omega^*\boldsymbol{l}^*$ for 100 trials. The ellipsoids in the figures are respectively defined by

$$(\Delta\boldsymbol{t}^*, V[\hat{\boldsymbol{t}}]^{-1}\Delta\boldsymbol{t}^*) = 1, \quad (\Delta\boldsymbol{\Omega}^*, V[\hat{\boldsymbol{R}}]^{-1}\Delta\boldsymbol{\Omega}^*) = 1, \quad (20)$$



(a) (b)

Fig. 5 Bootstrap errors (our method): (a) translation; (b) rotation.



(a) (b)

Fig. 6 Bootstrap errors (least squares): (a) translation; (b) rotation.

where $V[\hat{\boldsymbol{t}}]$ and $V[\hat{\boldsymbol{R}}]$ are computed by approximating $\bar{\boldsymbol{R}}$, $\{\bar{\boldsymbol{x}}_\alpha\}$, and ϵ^2 by $\hat{\boldsymbol{R}}$, $\{\hat{\boldsymbol{x}}_\alpha\}$, and $\hat{\epsilon}^2$ on the right-hand side of Eq. (8). These ellipsoids indicate the standard deviation of the errors in each orientation [6]. The cubes in the figures are displayed as a reference.

We compared our method with the naive least-squares method; we simply replaced the matrix $\boldsymbol{W}_\alpha$ by the unit matrix $\boldsymbol{I}$. Fig. 6 shows the result that corresponds to Fig. 5; the ellipsoids and the cubes are the same as in Fig. 5.

Comparing Figs. 5 and 6, we can confirm that our method improves the accuracy of the solution as compared with the least-squares method. We can also see that errors for our method distribute around the ellipsoids, indicating that our method already attains the theoretical accuracy bound; no further improvement is possible.

The above visual observation can be given quantitative measures. We define the *bootstrap standard deviations* by

$$S_t^* = \sqrt{\frac{1}{B}\sum_{b=1}^{B}\|\Delta\boldsymbol{t}_b^*\|^2},$$
$$S_R^* = \sqrt{\frac{1}{B}\sum_{b=1}^{B}(\Delta\Omega_b^*)^2}, \qquad (21)$$

where B is the number of bootstrap samples and the subscript b labels each sample. The corresponding standard deviations for the (estimated) theoretical lower bound are $\sqrt{\mathrm{tr}V[\hat{\boldsymbol{t}}]}$ and $\sqrt{\mathrm{tr}V[\hat{\boldsymbol{R}}]}$, respectively; these indicate the average radii of the ellipsoids defined by eqs.(20) and displayed in Figs. 5 and 6.

Table 1 Bootstrap standard deviations and the theoretical lower bounds.

	Translation	Rotation
Our method	0.16cm	0.16°
Least squares	0.25cm	0.45°
Lower bounds	0.16cm	0.15°

Fig. 7 A real image of a real building.

Table 1 lists the values of S_t^* and S_R^* for our method and the least-squares method ($B = 1000$) together with their theoretical lower bounds. We can see from this that our method is indeed superior to the least-squares method and that the accuracy of our solution is very close to the theoretical lower bound.

Fig. 7 is a real image of a real building for which a design plan is available. We manually input the feature points marked by white dots and used the noise model of Eq. (4). The computation converged after four iterations for thresholds $\epsilon_t = 0.1$cm and $\epsilon_R = 0.01°$. Fig. 8 displays the building graphically reconstructed from the design plan viewed from above together with the estimated camera coordinate axes; the ellipse in the figure indicates the ellipsoid corresponding to those in Figs. 5(a) and 6(a) enlarged by three times. We also evaluated the reliability of the solution by both theoretical analysis and bootstrap. Table 2 is the result corresponding to Table 1. We can again confirm that our method is superior to the least-squares method and that our method almost attains the theoretical bound, which can be viewed as describing the probability distribution of the estimated location.

6. Concluding Remarks

We have discussed optimal estimation of the current location of a mobile robot by matching an image of the scene taken by the robot with the model of the environment. We have first presented a theoretical accuracy bound defined independently of solution techniques and then given a method that attains it; our method is truly "optimal" in that sense. Using real images, we have demonstrated that our method is su-

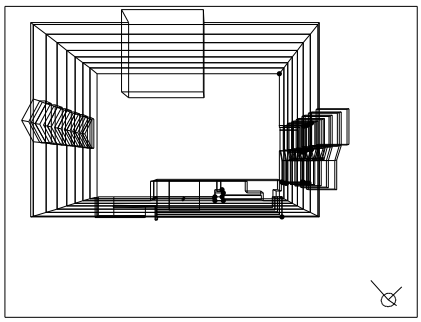


Fig. 8 Estimated current location and its reliability.

Table 2 Bootstrap standard deviations and the theoretical lower bounds.

	Translation	Rotation
Our method	44.1cm	1.31°
Least squares	47.3cm	1.39°
Lower bounds	43.7cm	1.29°

perior to the naive least-squares method. We have also confirmed the theoretical predictions of our theory by applying the bootstrap procedure.

In our method, we need not know the absolute magnitude of image noise; it is estimated a posteriori. If it is known in advance, however, we can make use of the redundancy of that knowledge to do various types of verification of our assumptions. For example, we should question the correctness of the feature matching if the estimated noise level is very large as compared with the known value.

The aim of this paper is not to propose a novel technique for robot self-localization but to discuss the statistical aspects, in particular the theoretical accuracy bound and optimality of the solution. The analysis given here can be applied to not only mobile robots but many other problems including robot arm manipulations and a posteriori human analyses in various circumstances.

Acknowledgments

The authors thank Hiroyuki Tsuzuki of OKI Data Systems for doing system implementation while he was at the graduate school of Gunma University.

References

[1] N. Ayache and O. D. Faugeras, "Building, registrating, and fusing noisy visual maps," *Int. J. Robotics Research*, vol. 7, no. 6, pp. 45–65, 1988.

[2] M. Betke and L. Gurvits, "Mobile robot localization using landmarks," *IEEE Trans. Robotics Automation*, vol. 13, no. 2, pp. 251–263, 1997.

[3] B. Efron and R. J. Tibshirani, *An Introduction to Bootstrap*, Chapman-Hall, New York, 1993.

[4] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Comm. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[5] K. Kanatani, "Renormalization for motion analysis: Statistically optimal algorithm," *IEICE Trans. Inf. & Syst.*, vol. E77-D, no. 11, pp. 1233–1239, 1994.

[6] K. Kanatani, *Statistical Optimization for Geometric Computation: Theory and Practice*, Elsevier, Amsterdam 1996.

[7] K. Sugihara, "Some location problems for robot navigation using a single camera," *Comput. Vis. Gr. Image Process.*, vol. 42, pp. 112–129, 1988.

[8] R. E. Suorsa and B. Sridhar, "A parallel implementation of a multisensor feature-based range-estimation method," *IEEE Trans. Robotics Automation*, vol. 10, no. 6, pp. 755–768, 1994.

[9] K. T. Sutherland and W. B. Thompson, "Localizing in unconstrained environment: Dealing with the errors," *IEEE Trans. Robotics Automation*, vol. 10, no. 6, pp. 740–754, 1994.

[10] R. Talluri and J. K. Aggarwal, "Mobile robot self-location using model-image feature correspondence," *IEEE Trans. Robotics Automation*, vol. 12, no. 1, pp. 63–77, 1996.

[11] Y. Yagi, Y. Nishimitsu and M. Yachida, "Map-based navigation for a mobile robot with ominidirectional image sensor COPIS," *IEEE Trans. Robotics Automation*, vol. 11, no. 5, pp. 634–648, 1995.

Kenichi Kanatani received his Ph.D. in applied mathematics from the University of Tokyo in 1979. He is currently Professor of computer science at Gunma University. He is the author of *Group-Theoretical Methods in Image Understanding* (Springer, 1990), *Geometric Computation for Machine Vision* (Oxford University Press, 1993) and *Statistical Optimization for Geometric Computation: Theory and Practice* (Elsevier, 1996).

Naoya Ohta received his ME from the Tokyo Institute of Technology in 1985 and his Ph.D. from the University of Tokyo in 1998. He engaged in research and development of image processing systems at the Pattern Recognition Research Laboratories of NEC. He is currently Assistant Professor of computer science at Gunma University. He was a research affiliate of the Media Laboratory in MIT from 1991 to 1992.