

SIGVerse を用いた自己位置と位置概念の同時推定に関する研究

Research on Simultaneous Estimation of Self-Location and Location Concepts using SIGVerse

谷口 彰^{*1}

Akira Taniguchi

吉崎 陽紀^{*1}

Haruki Yoshizaki

稻邑 哲也^{*2}

Tetsunari Inamura

谷口 忠大^{*1}

Tadahiro Taniguchi

^{*1}立命館大学

Ritsumeikan University

^{*2}国立情報学研究所

National Institute of Informatics

In this paper, we proposed a novel learning method which can estimate self-location of a robot and concepts of location simultaneously. A robot performs a probabilistic self-localization from sensor data. We integrate ambiguous speech recognition results with it based on Bayesian approach. Experimental results show that a robot can obtain words for several locations and make use of them in self-localization task using SIGVerse.

1. はじめに

近年，サービスロボット等の人間の生活環境で動作するロボットの実現が求められている。こうしたロボットは，様々な生活環境において周囲の環境を認知し，さらに人とのインタラクションを通して知識を学習していくことが重要であると言える。また，ロボットはセンサを通じて環境を認識する際，物理世界に存在する多くの不確実性への対処が重要である[1]。

本稿では，場所に関する知識獲得に注目し，環境を自己位置推定を行いながら移動するロボットに，人が場所の名前を教示することで場所に対応した語彙を獲得させることを目標とする。ロボットは自己位置推定を行う際，センサ情報から確率的な自己位置推定を行う。本稿では，これに不確実な音声認識結果を統合することを行う。すなわち本稿では，不確実な音声認識と自己位置情報を相互に有効活用した，自己位置と語彙の同時推定モデルを提案する。また，これを SIGVerse 上に実装し検証を行う。SIGVerse とは，サーバクライアント形式の社会的知能発生学シミュレータである。SIGVerse は 3 次元空間上で，力学計算，知覚シミュレーション，コミュニケーションのシミュレーションを同時に使うことができる[2]。

山田ら[3]の研究では，多様な言い回しの教示から，指示対象のカテゴリとそれを表す音素系列を同時に学習する手法が提案されている。学習後，学習に用いていない位置情報を入力し，場所に対応した正しいキーワードの出力を行っているが，本研究での提案モデルは，獲得した知識からロボット自身の位置情報へのフィードバックを行うものである。

2. 自己位置と位置概念の推定モデル

本研究では，MCL (Monte Carlo Localization) を自己位置推定の手法として利用し，MCL に位置概念を導入し拡張したモデルを提案する。位置概念とは，場所に対応した語彙のことである。これには場所の名前とその名前によって表現される場所の位置分布という二つの知識が含まれるものとする。

地図のある環境上で移動ロボットに建物内を移動させ，自己位置推定を行わせることを想定する。ロボットが学習対象の場所に移動したとき，人が場所の名前の教示を複数回行うことで，位置概念を学習させる。提案手法の全体像を表す概略図を図 1 に示す。

連絡先: 谷口彰，立命館大学情報理工学研究科，
a.taniguchi@em.ci.ritsumei.ac.jp

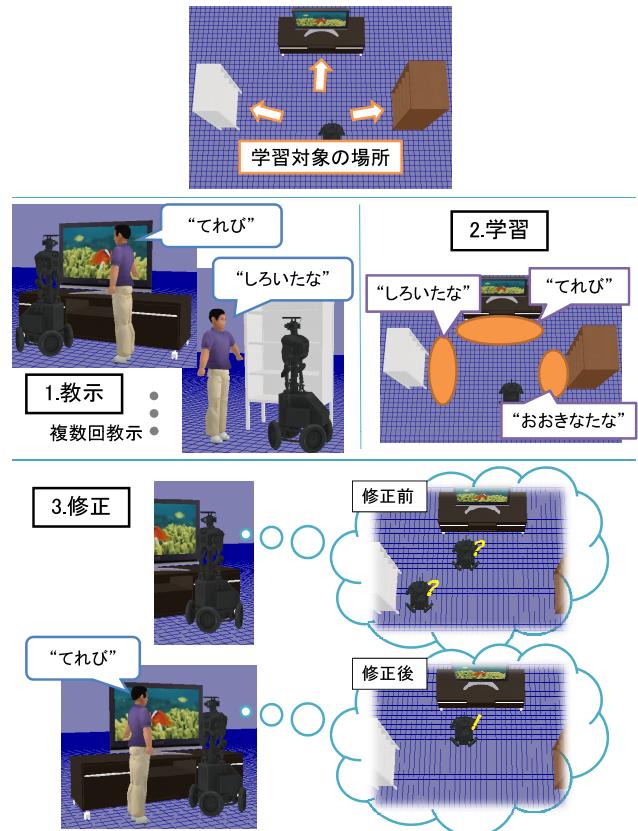


図 1: 提案手法の全体像を表す概略図

ロボットには事前に単語の知識を持たせず，日本語の音節のみを認識できるものとする。

本稿では，場所の名前を複数回教示されたときの位置概念の学習方法と，位置概念を獲得したロボットが場所の名前を聞いたときの自己位置情報の修正について考える。

2.1 位置概念の定義

学習により獲得した位置概念の集合 Λ を，式 (1) の様に表す。 L は位置概念の数である。 i 番目の位置概念は，式 (2) の

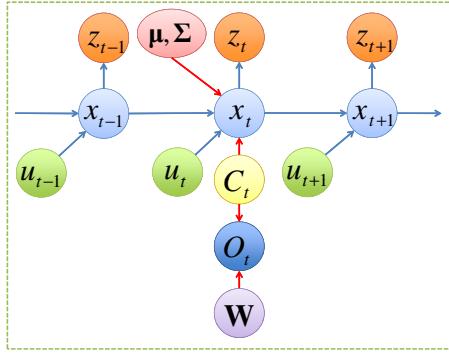


図 2: 位置概念 MCL のグラフィカルモデル

様に定義する .

$$\lambda = \{\lambda_1, \lambda_2, \dots, \lambda_L\} \quad (1)$$

$$\lambda_i = \{W_i, \mu_i, \Sigma_i\} \quad (2)$$

このとき , W_i は学習により得た i 番目の場所の名前である . 式 (3) の l_j^i は W_i の j 番目の文字であり , d_i は W_i の文字数である . 場所の名前の集合は , 式 (4) の様に表せる . また , 場所の位置分布はガウス分布を仮定する . そのため , μ_i は i 番目の分布の平均 , Σ_i は i 番目の分布の共分散行列である . 平均の集合は , 式 (5) , 共分散行列の集合は , 式 (6) の様に表せる .

$$W_i = (l_1^i l_2^i \cdots l_{d_i}^i) \quad (3)$$

$$\mathbf{W} = \{W_1, W_2, \dots, W_L\} \quad (4)$$

$$\boldsymbol{\mu} = \{\mu_1, \mu_2, \dots, \mu_L\} \quad (5)$$

$$\boldsymbol{\Sigma} = \{\Sigma_1, \Sigma_2, \dots, \Sigma_L\} \quad (6)$$

時刻 t における , ロボットが聞き取った音声情報 O_t に対応した位置概念の指標を C_t で表す . C_t は , 位置概念の指標の集合を \mathbf{C} としたとき , 式 (7) で表す様に 1 から L までの離散値をとる .

$$C_t \in \mathbf{C} = \{1, 2, \dots, L\} \quad (7)$$

つまり , C_t 番目の位置概念は , $\lambda_{C_t} = \{W_{C_t}, \mu_{C_t}, \Sigma_{C_t}\}$ と表せ , W_{C_t} μ_{C_t} Σ_{C_t} は , それぞれ C_t 番目の位置概念の場所の名前 , 位置分布の平均 , 共分散行列を表す .

2.2 位置概念を導入した自己位置推定

MCL に位置概念を導入したグラフィカルモデルを図 2 に示す . 青い矢印で示した部分が通常の MCL , 赤の矢印で示した部分が提案手法による拡張部分である . グラフィカルモデルの各要素についてまとめたものを表 1 に示す .

MCL の導出式に O_t を加えた式を , 式 (8) に示す .

$$\begin{aligned} & p(x_{0:t} | z_{1:t}, u_{1:t}, O_{1:t}) \\ & \propto p(z_t | x_t) p(O_t | x_t) p(x_t | x_{t-1}, u_t) \\ & \quad \cdot p(x_{0:t-1} | z_{1:t-1}, u_{1:t-1}, O_{1:t-1}) \end{aligned} \quad (8)$$

式 (8) で導出された $p(O_t | x_t)$ について , グラフィカルモデルの関係により導出したものを , 式 (9) に示す . この式は , x_t という場所で O_t という言葉を聞き取ったときの確率を表しており , 計測モデル $p(z_t | x_t)$ と同じく尤度として計算を行う .

$$p(O_t | x_t) \propto \sum_{C_t} p(O_t | \mathbf{W}, C_t) p(x_t | \boldsymbol{\mu}, \boldsymbol{\Sigma}, C_t) p(C_t) \quad (9)$$

表 1: グラフィカルモデルの各要素表

x_t	ロボットの自己位置
u_t	制御値
z_t	計測値
C_t	位置概念の指標
O_t	音声認識した単語
\mathbf{W}	学習した場所の名前
$\boldsymbol{\mu}, \boldsymbol{\Sigma}$	場所の位置分布 (平均 , 共分散行列)

また本研究では , $p(C_t)$ は無情報を仮定し , $p(C_t) = \frac{1}{L}$ の様に一様分布として近似する .

$p(O_t | \mathbf{W}, C_t)$ については , 編集距離によって場所の名前と認識文字の近さを表す確率の式を , 式 (10) に定義する . β は編集距離の値による影響度合いを確率的に反映するパラメータであり , LD は編集距離を計算する関数である .

編集距離とは , 二つの文字列の異なり具合を量量化するものであり , 文字の挿入・削除・置換により一方を他方に変形するまでの最小手順回数を数えたものである . 本研究では , 音声認識器による認識単語や獲得した場所の名前の単語の近さを編集距離によって測る .

$$p(O_t | \mathbf{W}, C_t) \propto \exp(-\beta \text{LD}(O_t, W_{C_t})) \quad (10)$$

$p(x_t | \boldsymbol{\mu}, \boldsymbol{\Sigma}, C_t)$ については , 場所の位置分布と自己位置情報の近さを表す確率の式を , 式 (11) に定義する . この式は , 次元数 D の多変量ガウス分布を表している .

$$\begin{aligned} & p(x_t | \boldsymbol{\mu}, \boldsymbol{\Sigma}, C_t) \\ & \propto \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma_{C_t}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x_t - \boldsymbol{\mu}_{C_t})^T \boldsymbol{\Sigma}_{C_t}^{-1} (x_t - \boldsymbol{\mu}_{C_t})\right) \end{aligned} \quad (11)$$

2.3 位置概念の学習

発話場所と認識単語による複数の学習用データから場所の名前と場所の位置分布をギブスサンプリングによって推定する . 位置概念の数 L は既知とする . ギブスサンプリングによる位置概念学習の流れを表したイメージ図を図 3 に示す . この図の場合 , 三つの物体前の場所を学習対象としている .

各ステップについて説明する . 詳細な式については , 文献 [4] を参照のこと .

1. $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ を初期化

初期の位置分布は全て , $\boldsymbol{\mu}_c = (\text{一定の範囲内に一様乱数})$ $\boldsymbol{\Sigma}_c = \begin{bmatrix} \sigma_{initial} & 0 \\ 0 & \sigma_{initial} \end{bmatrix}$ とする .

2. \mathbf{W} の情報無しで , C_t をサンプリング

どのデータがどの位置概念の指標を表すかを推測する . このとき , 教示された時刻 t の集合を $T_o = \{t_1, t_2, \dots, t_N\}$ とする . N は教示データ数である .

$$C_t \sim p(x_t | \boldsymbol{\mu}_{C_t}, \boldsymbol{\Sigma}_{C_t}) p(O_t | W_{C_t}) \quad (12)$$

C_t のサンプリングは , $C_t \in \mathbf{C} = \{1, 2, \dots, L\}$ についてそれぞれ確率を出し , 亂数により選択する . ただし , ステップ 2 では \mathbf{W} の情報が無いため , 式 (13) を計算し , サンプリングする .

$$C_t \sim p(x_t | \boldsymbol{\mu}_{C_t}, \boldsymbol{\Sigma}_{C_t}) \quad (13)$$

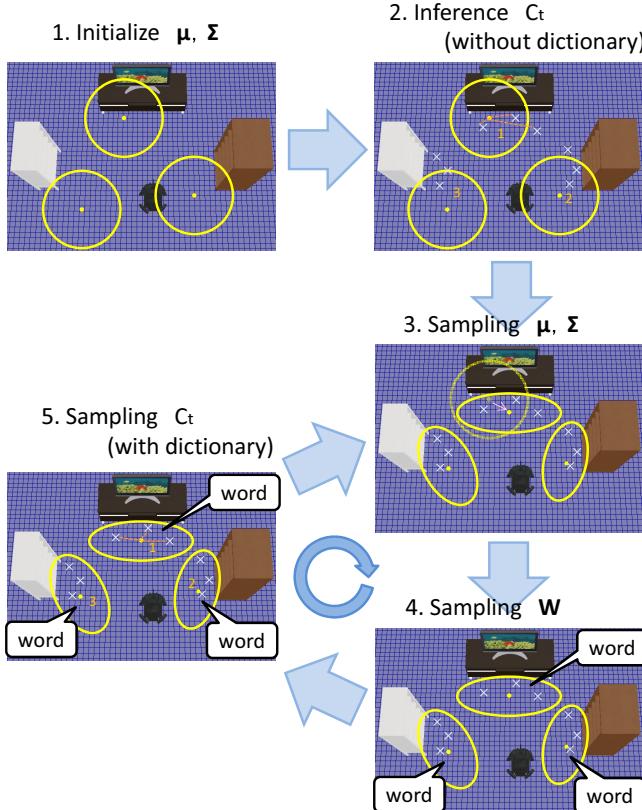


図 3: ギブスサンプリングによる位置概念学習の流れ

3. μ, Σ をサンプリング

場所の位置分布は、式 (14) の様に $c \in C = \{1, 2, \dots, L\}$ についてそれぞれ計算し、サンプリングする。

$$\mu_c, \Sigma_c \sim \prod_{\substack{c=C_t \\ t \in T_o}} \left(p(x_t | \mu_c, \Sigma_c) \right) p(\mu_c, \Sigma_c) \quad (14)$$

これは、学習データ ($t \in T_o$) の中で $C_t = c$ であるデータのみを選び、 $p(\mu_c, \Sigma_c)$ に掛け合わせたものになっている。式 (15) の様に、事前分布 $p(\mu_c, \Sigma_c)$ をガウス・ウィシャート分布とし、ガウス分布に対するベイズ推論 [5] を行うことで事後分布を計算する。 Λ は精度行列であり、共分散行列の逆行列である ($\Lambda = \Sigma^{-1}$)。 m_0, β_0, V_0, ν_0 は、事前分布のハイパーパラメータを表す。

$$p(\mu, \Lambda) = \mathcal{N}(\mu | m_0, (\beta_0 \Lambda)^{-1}) \mathcal{W}(\Lambda | V_0, \nu_0) \quad (15)$$

μ, Λ のサンプリングは、式 (16)、式 (17) の様に表せる。

$$\Lambda \sim \mathcal{W}(\Lambda | V_N, \nu_N) \quad (16)$$

$$\mu \sim \mathcal{N}(\mu | m_N, (\beta_N \Lambda)^{-1}) \quad (17)$$

学習データ中に該当する位置概念が存在しない場合、パラメータ m_0 を一定の範囲内となる一様乱数により与える。

4. W をサンプリング

場所の名前は、式 (18) の様に $c \in C = \{1, 2, \dots, L\}$ に

発話単語	認識単語				
	てりまえ	てびまえ	てぎくなえ	てれなえ	てりまえ
てれびまえ	てりまえ	てびまえ	てぎくなえ	てれなえ	てりまえ
	てれなえ	てびまえ	てぎくなえ	ててぎなえ	てりびなえ
しろいたな	しろいたの	しにたな	しろいたが	しのたが	しめたな
	ちにたの	つたな	しのにたな	しのたな	しるいたな
つくえ	つくべ	つくね	つくう	つくえ	つくめ
	つくね	つくけ	つつね	つくね	つくめ
ごみばこ	ごみわこ	ぐびわこ	ぐみわこ	ぐびわこ	くみわこ
	ごみわこ	ぐみわこ	ぐみわこ	ぐみばこ	ぐみばく

図 4: 日本語音節単語辞書を用いた Julius の認識結果

についてそれぞれ計算し、サンプリングする。

$$W_c \sim \prod_{\substack{c=C_t \\ t \in T_o}} \left(p(O_t | W_c) \right) p(W_c) \quad (18)$$

このとき、 $p(W_c)$ は、無情報を仮定し、 $p(W_c) = \frac{1}{N}$ の様に一様分布として近似する。

W_c のサンプリングは、 $t \in T_o = \{t_1, t_2, \dots, t_N\}$ についてそれぞれ確率を出し、乱数により選択する。学習データの中に該当する位置概念がない場合、 $p(W_c)$ が一様分布より、全ての学習用データ単語から等確率で選択する。

5. W の情報有りで、 C_t をサンプリング
 W が推定されているため、式 (12) を計算し、 C_t をサンプリングする。サンプリング方法は、ステップ 2 と同様である。
6. ステップ 3~5 の過程を複数回繰り返す。

3. SIGVerse 上での実験

音声認識器には、大語彙連続音声認識システム Julius[6]^{*1}を利用した。Julius の単語辞書は、既存の大量語が登録された辞書を用いず、日本語音節のみを登録した単語辞書を使用する。これによって、発話音声を音節毎に分け音節列として認識することができる様になる。マイクには、SHURE 社の PG27 USB を使用し実験を行った。

3.1 位置概念の学習

3.1.1 実験条件

本実験では、位置概念の数は $L = 4$ とし、 $\sigma_{initial} = 10000$ 、ステップ 3 における事前分布のハイパーパラメータの初期値は、 $\beta_0 = 0.001, m_0 = [0, 0]^T, V_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \nu_0 = 1$ 、ステップ 6 での繰り返し回数は、10 回とした。発話場所は、四つの物体前付近とし、それぞれ 10 回ずつ合計 40 回分の発話データを与えた。発話単語はそれぞれ、“しろいたな”、“つくえ”、“ごみばこ”、“てれびまえ”とし、学習用データには、図 4 の各発話単語に対する認識結果を利用した。

3.1.2 実験結果

この設定で試行した学習結果から、1 例を図示したものを図 5 に示す。各位置分布の場所を円で囲んでいる。 $W_1 \sim W_4$ は、それぞれの分布に対応した場所の名前である。この結果から、四つの物体前の場所に対応した四つの位置概念が学習できていることが分かる。

*1 インストールバージョン : dictation-kit-v4.2-win .

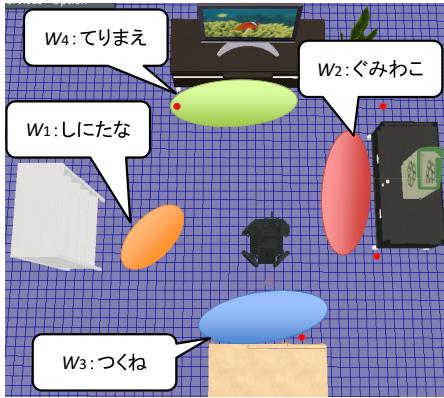


図 5: 学習結果例 1

3.2 学習した位置概念の評価のための比較実験

3.2.1 実験条件

次に、先の位置概念の学習結果を用いた自己位置情報の修正について確認した。SIGVerse 上でランドマーク型の MCL を実装し、通常の MCL と提案手法によって位置概念を学習した状態の MCL とで自己位置の推定誤差精度の比較を行った。空間上に四つのランドマークを図 5 の赤点で示した部分に設置した。ロボットのカメラ視野角内にランドマークがある場合、ランドマークを認識する。センサ情報として、認識したランドマークまでの距離と角度を測り、尤度計算を行う。カメラはロボット前方に設置しており、カメラ視野角は 45 度である。パーティクル数は $M = 1000$ とし、初期パーティクルは一様に分布した。ロボットは 1 step 毎に 1 回の制御動作を行う。ロボットは 400 step 移動し、各物体前の場所を“しろいたな”、“つくね”、“ごみばこ”、“てれびまえ”と順番に巡る様な動作をする。この時、全ての試行において、ロボットは同じ動作とした。提案手法でのみ、各物体前付近の場所に来たとき場所の名前を教示した。教示単語は、図 4 の各発話単語に対する認識結果からランダムに選ぶ。

推定誤差の評価には、上田ら [7] の評価方法を参考にした。まず、空間の床部分を xy 平面とし、step 每に、式 (19) の様に xy 平面での推定誤差を記録する。

$$e_t = \sqrt{(\bar{x}_t - x_t^*)^2 + (\bar{y}_t - y_t^*)^2} \quad (19)$$

このとき、 $\bar{x}_t = \sum_i^M w_t^{(i)} x_t^{(i)}$, $\bar{y}_t = \sum_i^M w_t^{(i)} y_t^{(i)}$ であり、 x_t^*, y_t^* はロボットの真の位置座標である。試行後、 e_t の平均値 \bar{e}_r と、区間 $[0, \gamma]$ が e_t の 95% 以上を含む γ の最小値 e_r^T を計算する。 e_r^T は推定の安定性を示す指標とする。

3.2.2 実験結果

10 回分の実験結果を \bar{e}_r , e_r^T によって比較したグラフを、図 6 に示す。どちらの指標でも、推定誤差は提案手法の方が小さく、この結果で有意水準 5% の片側 t 検定を行ったところ、 \bar{e}_r , e_r^T 共に有意差が見られた。よって、自己位置情報の修正を行えるような位置概念が獲得できていることが確認できた。

4.まとめと考察

本研究では、移動ロボットによる場所に対応した語彙獲得を目的とし、位置概念を獲得と自己位置情報を修正する方法を検討し、SIGVerse 上での実験を行った。位置概念の学習では、多くの場合で四つの物体前の場所に位置概念がそれぞれ形

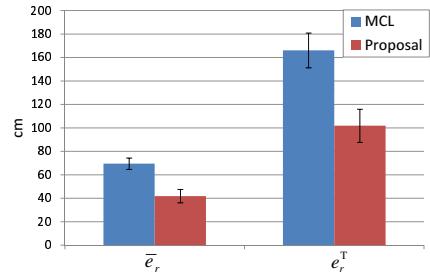


図 6: MCL と提案手法による推定誤差の比較

成されたが、二つの物体前の場所を一つの位置概念が包含するような位置分布が形成された場合も見られた。

本稿では、自己位置が既知の場合での位置概念の学習を行つたが、今後は位置概念を獲得していない状態から、自己位置推定の情報と与えられた言語情報を用いた自己位置情報の修正と位置概念の学習を行う方法について検討していきたい。また本研究では、地図情報が既知である MCL を仕様したモデルであったが、地図情報が与えられていない SLAM (Simultaneous Localization And Mapping) を使用したモデルへの拡張が考えられる。

参考文献

- [1] S. Thrun, W. Burgard, D. Fox, 上田隆一 (訳). 確率ロボティクス. 毎日コミュニケーションズ, 2007.
- [2] Tetsunari Inamura et al. Simulator platform that enables social interaction simulation –SIGVerse: SocioIntelliGenesis simulator–. In *IEEE/SICE International Symposium on System Integration*, pp. 212–217, 2010.
- [3] 山田雄治, 服部公央亮, 田口亮, 梅崎太造, 保黒政大, 岩橋直人, 船越孝太郎, 中野幹生. 連続音声から場所の名前を学習する自律移動ロボット. 一般社団法人情報処理学会 全国大会講演論文集, Vol. 2011, No. 1, pp. 237–239, mar 2011.
- [4] 谷口彰, 吉崎陽紀, 谷口忠大. 不確実な音声認識を前提とした自己位置と語彙の同時推定モデル. 第 57 回システム制御情報学会研究発表講演会, 2013.
- [5] C.M. ビショップ著 元田 浩・栗田多喜夫 監訳 横口知之他 監訳. パターン認識と機械学習 上 ベイズ理論による統計的予測. 丸善出版, 2007.
- [6] 荒木雅弘. フリーソフトでつくる音声認識システム: パターン認識・機械学習の初步から対話システムまで. 森北出版株式会社, 2007.
- [7] 上田隆一, 新井民夫, 浅沼和範, 梅田和昇, 大隅久. パーティクルフィルタを利用した自己位置推定に生じる致命的な推定誤りからの回復法. 日本ロボット学会誌, Vol. 23, No. 4, pp. 466–473, 2005.