

観測空間ダイナミクスの外延を用いた 不完全な知覚を有するロボットの行動生成

Motion generation of robot with incomplete perception
using extrapolation of dynamics in observation space

小林 祐一^{*1*2}
Yuichi Kobayashi

栗田英介^{*1}
Eisuke Kurita

郷古学^{*3}
Manabu Gouko

^{*1}東京農工大学

Tokyo University of Agriculture and Technology

^{*2}理化学研究所 RTC

RIKEN RTC

^{*3}東北学院大学

Tohoku Gakuin University

Sensors of robots that act in unstructured environment sometimes do not provide complete observation, due to occlusion or limitation of sensing range. This paper presents a motion generation method for robot with multiple sensors with limited sensing ranges. The proposed method introduce extension of the action-observation mapping to outside of the sensing range of a sensor, based on the diffusion-based learning of Jacobian matrices between control input and observation variable. Multiple observation spaces can be integrated by finding correspondence between the virtual observation spaces. When a target observation variable is given to the robot, it can generate a motion from an observation space toward the target with another observation space using the extended observation space. The proposed framework is verified by reaching motion toward the floor with a manipulator. Observation space by camera with limited view was extended and appropriate motion trajectories were obtained.

1. はじめに

ロボット・環境のモデルを事前に設定・解析することなくロボットの試行錯誤により適切な行動を生成させるアプローチは、強化学習のロボットへの適用や認知発達ロボティクス分野 [Asada 01] などで活発に研究されている。一方、ロボットの活動範囲をより多様な領域に広げる際に生じる問題として、センサの計測範囲の限界があげられる。例えば、カメラから得られる視覚情報には視野範囲という制限があり、距離センサはある一定の距離を超えると物体までの距離計測が行えなくなる。また、視覚・距離センサ共通の問題として遮蔽により物体の奥側の領域の計測は行えないという問題がある。

複数のセンサ情報を統合することでより高い計測性能を得ることを目指した研究として、センサフュージョン [Luo 02] があげられる。センサフュージョンのアプローチでは、センサによる計測に誤差が含まれるという意味での不完全さを扱うことができる [Thrun 05] が、センサの観測が全く得られない状況を陽に扱った研究はみられない。移動ロボットによる自己位置同定問題では、センサの観測が得られない状況においてデッドレコニング (dead reckoning) を行うことでオドメトリの情報から自己位置を推定することができるが、このような世界座標系でのモデルを用いずに望ましい行動を生成できることが望ましい。

一方、事前知識を用いずにロボットの観測と行動、報酬情報のみから行動を生成する枠組みとして強化学習が知られる。センサによる状態観測の不完全性 (観測誤差, 知覚騙し: 実際の状態は異なるのに同じセンサ観測が得られる, など) は部分観測マルコフ決定過程 (Partially observable Markov decision process: POMDP) [Kaelbling 98] として議論されている (例えば [McCallum 95, 木村 96])。強化学習の問題定式化は非常に一般的で、広い汎用性を持つが、その反面、観測・行動の履歴を含めた強化学習の枠組みは多くの試行錯誤を必要とすることが知られている。

これに対して本研究では、観測範囲に制限のあるセンサを複

数搭載したロボットのための行動生成法を提案する。観測誤差 (ノイズ) の意味での不完全性ではなく、観測範囲外では全くセンサ信号が得られないような問題に対して、センサ (観測) 変数と制御入力の情報にのみ立脚した非線形の状態推定にもとづいた行動生成の方法を提案する。また、POMDP での定式化よりも狭いクラスの問題として、センサの観測および状態遷移が制御入力に対して微分可能な連続関数として表現できる問題を考える。この連続性を利用し、センサによる観測範囲内から外にでもどって来たときの「整合性」にもとづいた観測空間の外延を行うことで、観測範囲の近傍における状態推定を可能にする方法を提案する。連続性を利用する方法として、Luo *et al.* による拡散学習 [Luo 98] を用いる。仮想的にセンサの観測空間を拡大し、本来重なり合わない 2 つのセンサ観測空間を仮想的に重ねることで、現在は観測範囲外であるもう一方のセンサ観測空間上の点への行動生成を実現する。

提案手法の強化学習に対する利点としては、観測と状態遷移の連続性という制約を加え、観測可能範囲の周辺に限定することで、より簡単な学習則での状態推定が実現可能である点があげられる。また、報酬というタスクに依存した情報を用いていないことから、異なるタスクが与えられたときに共通して利用可能な情報を獲得している (異なるタスクに対してゼロから再学習する必要がない) ことがあげられる。

2. 問題設定

本論文では、2次元状態空間と2次元の観測変数で議論を行うが、3次元以上の空間に対しても原理的に対応可能である。ロボットと環境からなる系の状態空間を \mathcal{X} とし、状態変数を $x \in \mathcal{X} \subset \mathbb{R}^2$ とする。ロボットに搭載した複数のセンサによる観測を $s^{(1)} \in \mathcal{S}^{(1)}, \dots, s^{(h)} \in \mathcal{S}^{(h)}$ とする。ここで、センサ i の観測変数空間を $\mathcal{S}^{(i)} \subset \mathbb{R}^2$ 、ロボットに搭載したセンサの総数を h とする。ロボットの制御入力を $u \in \mathbb{R}^2$ とする。

状態変化 \dot{x} と制御入力 u の関係は次式のように表されるものとする。

$$\dot{x} = F(x)u, \quad F(x) \in \mathbb{R}^{2 \times 2} \quad (1)$$

$F(x)$ は 1 階微分可能な関数であり、ロボットには $F(x)$ を事

連絡先: 小林祐一, 東京農工大学, 東京都小金井市中町 2-24-16, yu-koba@cc.tuat.ac.jp

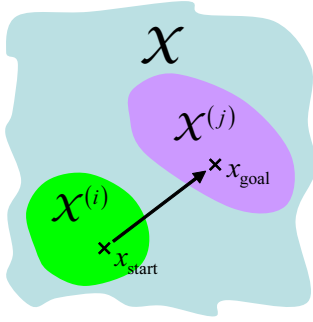


図 1: Motion generation from a observation space to another

前知識として与えない．また，観測 $s^{(i)}$ にはそれぞれ観測可能な範囲が存在し，範囲外の領域では適切な観測が得られない．この際，ロボットは各センサ i について現在観測範囲内であるかどうかを判別できるものとする．センサ i が観測範囲内であるロボットの状態空間を $\mathcal{X}^{(i)}$ とする (Fig.1 参照)． $\mathcal{X}^{(i)}$ から $S^{(i)}$ への写像は全単射であり，1 階微分可能であるとする．また，本研究ではノイズによる観測の不完全性は考慮しない．

ロボットのタスクは Fig.1 に示すように，初期状態 x_{start} から目標状態 x_{goal} に遷移させることである．ここで，次式の関係が成り立つものとする．

$$x_{\text{start}} \in \mathcal{X}^{(i)}, \quad x_{\text{start}} \notin \mathcal{X}^{(j)} \quad (2)$$

$$x_{\text{goal}} \notin \mathcal{X}^{(i)}, \quad x_{\text{goal}} \in \mathcal{X}^{(j)} \quad (3)$$

ただし，ロボットは状態変数を観測できないため，実際には x_{start} に対応するセンサ i の観測値 $s_{\text{start}}^{(i)}$ ， x_{goal} に対応するセンサ j の観測値 $s_{\text{goal}}^{(j)}$ が与えられる．このタスクを達成するためには，Fig.1 中の $\mathcal{X}^{(i)}$ ， $\mathcal{X}^{(j)}$ いずれにも属さない状態を経由した経路生成と動作生成を行う必要がある．

3. 不完全知覚ロボットの行動生成法

ロボットは，(1) の x と u の関係を事前知識として持たないため，ロボット自ら動き回ることによって得られる u と $s^{(i)}$ から， $s^{(i)}$ の変化と u の間の関係を Jacobi 行列として獲得する．しかし，各センサには観測範囲の制限が存在し，観測範囲外での状態では適切な観測が得られない．そこで，センサ i について，その観測範囲内のある状態 $x \in \mathcal{X}^{(i)}$ から範囲外 $x' \notin \mathcal{X}^{(i)}$ に遷移させ，再び観測範囲内のある状態 $x'' \in \mathcal{X}^{(i)}$ に遷移するという動きを繰り返し行わせる．その際，観測範囲外に出る前に得た観測 $s^{(i)}$ およびその後の制御入力の時系列から推定される観測変数値と，観測範囲内に戻ってきたときに得た観測 $s^{(i)''}$ との間の整合性を評価することで仮想的な Jacobi 行列の推定を行う．これにより，観測範囲外での観測変数を仮想的に推定し，観測範囲外での観測変数の「仮想的な外延」を行う．外延により得られたセンサ i の (仮想的な) 観測変数空間とセンサ j の観測範囲内が重なるとき，両者の対応をとることで，センサ i の観測空間とセンサ j の観測空間の関係を表現することができる．これにより，動作生成を実現する．なお，3.2 節までは単一のセンサの観測についての議論であるため，センサの番号 i を省いて s と表すものとする．

3.1 観測範囲内の Jacobi 行列の獲得法

センサ観測変数のダイナミクスが関数 $G(s) \in \mathbb{R}^{2 \times 2}$ により $\dot{s} = G(s)u$ で与えられるとする．時刻 t での観測を s_t とする

と，ロボットに制御入力 u_t を微小時間 Δt の間与えた後の観測 $s_{t+\Delta t}$ は， G により以下のように表される．

$$s_{t+\Delta t} = s_t + \int_t^{t+\Delta t} G(s_t)u_t dt \quad (4)$$

$G(s)$ を J_s によって以下のように近似する．

$$s_{t+\Delta t} - s_t = \int_t^{t+\Delta t} G(s_t)u_t dt \simeq J_{s_t}u_t \Delta t \quad (5)$$

観測空間 S を離散化した各格子点上でこの Jacobi 行列 J_s を推定する．2 次元格子上の $a, b \in \mathbb{Z}$ 番目の格子点を (a, b) と表し， (a, b) に対応した Jacobi 行列を $J_{a,b} \in \mathbb{R}^{2 \times 2}$ と定義する．観測範囲内の Jacobi 行列 $J_{a,b}$ の推定は，次式に示す二乗誤差 E_1 を最小化することで行う．

$$E_1 = \|s_{t+\Delta t} - s_t - J_{a,b}u\|^2 \quad (6)$$

勾配法を用いて $J_{a,b}$ を次式により更新する．

$$J_{a,b} \leftarrow J_{a,b} - c\alpha_1(k) \frac{\partial E_1}{\partial J_{a,b}} \quad (7)$$

ここで c は学習率である． $\alpha_1(k)$ は離散観測値 (a, b) の Jacobi 行列の学習回数 k に応じたガウス関数 $\exp(-k/K)$ ， K は定数である．

3.2 観測範囲外の状態の Jacobi 行列の獲得法

観測範囲外の Jacobi 行列の学習には 3. 章冒頭で述べた観測の整合性を用いる．今，センサが観測可能な状態から観測範囲外の状態に出て，合計 q 回の行動 (制御入力の印加) を経て再びセンサが観測可能な状態に戻って来たときと仮定する．この時， q 回行動後のロボットが推定する観測 s_J は次式で与えられる．

$$s_J = s_0 + \sum_{p=1}^q J_{a_p, b_p} u_p \quad (8)$$

ここで， s_0 は観測範囲外に出る直前の観測値， u_p は p ステップ目の制御入力， a_p, b_p は p ステップ目において推定された観測変数に対応する格子の座標である．観測範囲外から範囲内に戻ってきたとき，再び観測できた観測 s_r と観測範囲外の Jacobi 行列を用いて仮想的に推定した観測 s_J が等しくなる必要がある．この意味での整合性を評価する二乗誤差として次式で定義される $E_2 = \|s_r - s_J\|^2$ を導入する．これを用いて，勾配法にもとづいて J_{a_p, b_p} の更新を次式で行う．

$$J_{a_p, b_p} \leftarrow J_{a_p, b_p} + c\alpha_1 \frac{\partial E_2}{\partial J_{a_p, b_p}} \quad (9)$$

このような方法で仮想的な観測値に対する Jacobi 行列を推定する方法の問題点として，整合性を満たす観測範囲外の観測値が無限に存在することが挙げられる．そこで，解を一意に定めるため，Luo *et al.* が提案した拡散学習 [Luo 98] を用いる．拡散学習は，冗長自由度を有するマニピュレータの逆運動学を問題と解く際，マニピュレータ手先位置の空間的变化に対する Jacobi 行列の変化を最小にするという拘束条件を与えることで逆運動学の解を一意に定める手法である．拡散学習の拘束を勾配法で更新式に直したものを次式に示す．

$$J_{a_p, b_p} \leftarrow \frac{J_{a_{p-1}, b_p} + J_{a_{p+1}, b_p} + J_{a_p, b_{p-1}} + J_{a_p, b_{p+1}}}{4} \quad (10)$$

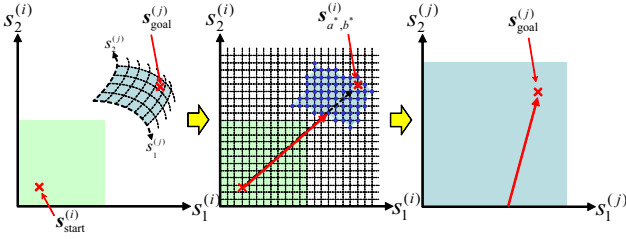


図 2: Integration of two observation spaces

これにより、解の連続性を拘束条件として用いて解を一意に定めることが出来る。また、観測範囲外での Jacobi 行列の情報が全く得られていない状況で (8) のような推定を行う際に初期解を与えることにも役立つ。(9) と (10) とを合わせ、次式のように Jacobi 行列の更新を行う。

$$J_{a_p, b_p} \leftarrow J_{a_p, b_p} + c\alpha_2(k) \frac{\partial E_2}{\partial J_{a_p, b_p}} + \alpha_1(k) \frac{J_{a_{p-1}, b_p} + J_{a_{p+1}, b_p} + J_{a_p, b_{p-1}} + J_{a_p, b_{p+1}}}{4} \quad (11)$$

ここで、 $\alpha_2(k) = 1 - \exp(-k/K)$ である。 $\alpha_1(k), \alpha_2(k)$ を加えることにより、繰り返し更新の初期段階では連続性の重みを大きく、後に整合性の重みを大きくしている。

3.3 二つの観測空間の統合による行動生成

Fig.2 に二つの観測空間の統合による行動生成の概略を示す。2種類のセンサ i および j を考え、センサ i の観測可能範囲から出発しセンサ j の観測可能範囲内のある点まで到達するような行動生成を考える。3.2 節で述べた観測範囲外への外延を用いることにより、次式を用いて (Jacobi 行列の獲得ができていない範囲で) センサ j の観測範囲内においてセンサ i の仮想観測変数値を得ることができる。

$$s_0^i + \sum_{p=1}^q J_{a_p, b_p}^{(i)} u_p \quad (12)$$

ただし、 s_0^i は観測範囲外に出る直前のセンサ i の観測値、 u_p は観測範囲外に出てから p ステップ目の制御入力であり、現在観測範囲外に出てから q ステップ行動を行っているとする。センサ i の格子 (a, b) における全センサの (仮想/実測) 観測変数を $s_{a,b}^{(i)} \in \mathbb{R}^{2 \times h}$, $a, b \in \mathbb{Z}$ と表す。 $s_{a,b}^{(i)}$ は次式で与えられる。

$$s_{a,b}^{(i)} = \left[s_{a,b}^{1(i)}, \dots, s_{a,b}^{h(i)} \right] \quad (13)$$

ただし、 $s_{a,b}^{k(i)} \in \mathbb{R}^2$ はセンサ i の観測空間における格子 (a, b) において得られるセンサ k の (仮想/実測) 観測変数を表す。

目標観測変数 $s_{\text{goal}}^{(j)}$ が与えられたとき、まず $s_{\text{goal}}^{(j)}$ に最も近いセンサ i の観測空間における格子点を以下のように求める。

$$(a^*, b^*) = \arg \min_{a,b} \| s_{\text{goal}}^{(j)} - s_{a,b}^{j(i)} \| \quad (14)$$

$s_{\text{goal}}^{(j)}$ に向かうための制御入力は次式で与えられる。

$$u = \begin{cases} \left(J_{a_p, b_p}^{(j)} \right)^{-1} \gamma \frac{s_{\text{goal}}^{(j)} - s_p^{(j)}}{\| s_{\text{goal}}^{(j)} - s_p^{(j)} \|} & \text{if } x_p \in \mathcal{X}^{(j)} \\ \left(J_{a_p, b_p}^{(i)} \right)^{-1} \gamma \frac{s_{a^*, b^*}^{(i)} - s_p^{(i)}}{\| s_{a^*, b^*}^{(i)} - s_p^{(i)} \|} & \text{otherwise,} \end{cases} \quad (15)$$

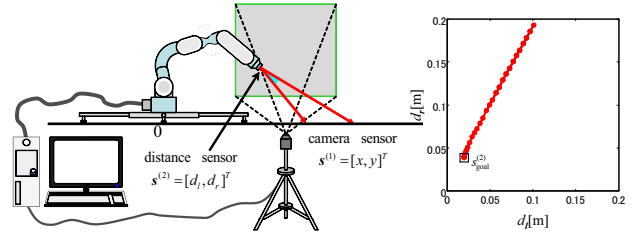


図 3: Manipulator with camera 図 4: Generated trajectory in sensor 2 space

ここで (a_p, b_p) は時間ステップ p における観測変数に最も近い格子点の番号、 $s_p^{(i)}$ はセンサ i の時間ステップ p における (仮想/実測) 観測変数値を表す。 γ はロボットの速度を決める定数である。動作生成の第 1 段階では (a^*, b^*) を目標格子点として (15) 下式のように制御を行う。ロボットがセンサ j の観測範囲内に入ると、センサ j に関する Jacobi 行列を用いて (15) 上式のように制御入力を決定する。

4. シミュレーション

提案行動生成法を、マニピュレータによる床へのリーチング動作において評価する。

4.1 マニピュレータのリーチングタスク

Fig.3 に示すような 2 リンクマニピュレータを考える。マニピュレータ手先位置をカメラにより検出することができるが、手先がカメラの視野範囲外に出ると観測値は得られなくなる。マニピュレータ手先には θ_s だけ計測方向の異なる距離センサが 2 個取り付けられている。距離センサは、センサから床までの距離に比例した値を、ある一定の計測限界距離まで与える。

2 つのセンサの観測可能範囲が重複していない (同時に 2 つのセンサが観測可能となることはない)。観測範囲内での Jacobi 行列の学習は観測範囲内をランダムに動きまわることによってデータを収集し、(7) の更新を用いて行う。1 格子点 (1 つの Jacobi 行列に相当する) あたり 1000 回の更新を行うことを学習の終了条件とする。また、学習に用いるパラメータは、 $c = 0.1$, $\gamma = 0.9$, $K = 100$ とする。

カメラから得られる手先位置情報を $s^{(1)} = [x, y]^T \in \mathcal{S}^{(1)}$ と表し、2 つの距離センサの値を $s^{(2)} = [d_l, d_r]^T \in \mathcal{S}^{(2)}$ と表す。マニピュレータへの制御入力は $u = [\dot{\theta}_1, \dot{\theta}_2]^T$ で定義される関節角速度である。リンク長はともに 0.6[m], θ_s は 30[deg], 距離センサの計測限界は 0.2[m] とする。カメラの視野範囲は $0.25 \leq x \leq 0.5$ [m], $0.25 \leq y \leq 0.5$ [m]^{*1} とする。この例題では、各センサ変数空間の格子は等間隔に設定し?節で述べた格子間隔の変更以外の提案手法の評価を行う。

4.2 拡散学習による観測変数空間の外延

まず、カメラの観測変数空間の外延の評価を行った。観測範囲の境界付近の点からスタートして観測範囲外に出て、14 ステップの行動の後に観測範囲内 (0.16[m] ほど下方) に戻ってくる動作を生成した。Fig.5 にそのときのアームの軌跡を示す。図左側の重なっている線は原点を始点とする 2 リンクアーム

*1 簡単のため、このシミュレーションではカメラにより得られる観測値が世界座標系での距離と等しいと仮定している。本提案手法は世界座標系での距離情報などを用いずにセンサから得られる情報のみから行動を生成することができるためこのような仮定は必要なく、シミュレーション結果をわかりやすくするためにこのように設定している。

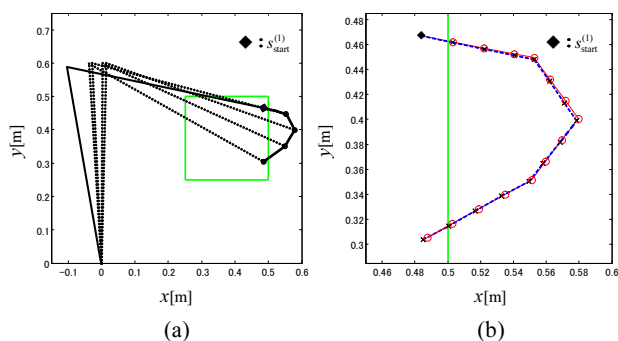


図 5: Estimated and actual hand position

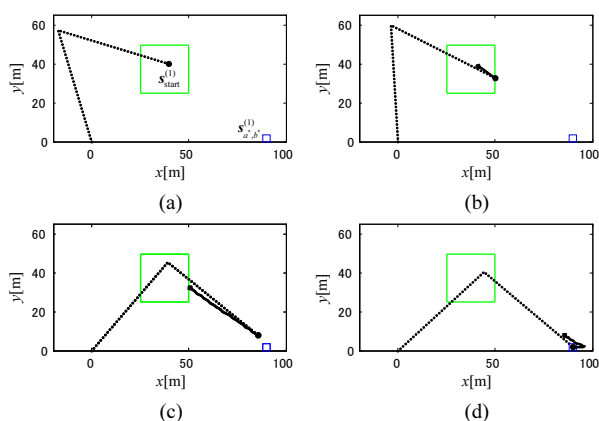


図 6: Generated trajectory of manipulator

ムを表している．正方形領域がカメラの観測範囲を表し，図 (a) 中の印は開始位置を示す．図右側は手先軌跡を拡大したものであり，×印の軌跡は提案手法により得られた推定値を表し，印の軌跡は実際の手先位置（観測範囲に制限がないとした場合の観測値）を表す．目標位置での観測値と推定値の間の誤差は 2.9×10^{-3} [m] であった．

4.3 2つの観測変数空間の統合による行動の生成

カメラにより観測される手先の初期位置は $s_{\text{start}}^{(1)} = [0.44, 0.3]^T$ ，目標とする距離センサの変数値を $s_{\text{goal}}^{(2)} = [0.02, 0.04]^T$ とする．(15)の制御則により実現されたマニピュレータの軌道を Fig.6, Fig.4 に示す．Fig.6 は $s^{(1)}$ で見た軌道の全体を表す．図中四角形外はカメラの観測範囲外であり，(d) は距離センサの観測範囲内の軌跡，(c) はどちらのセンサも観測範囲外となる軌跡を表している．Fig.4 は Fig.6(d) に対応する距離センサの観測範囲における $s^{(2)}$ での軌道を表す．全体の軌道から，両方のセンサが観測できない領域を通って目標センサ値を得られる状態まで遷移することができていることがわかる．また，カメラの観測空間では床付近で曲がった軌道を描いているが $s^{(2)}$ 空間内ではほぼ直線的な運動をしていることがわかる．これは，観測変数空間の間の写像が非線形であることが原因である．

4.4 考察

提案した動作生成法では，Jacobi 行列の情報を用いて各観測空間内で直線的に目標へ向かう経路生成のみを行っている．しかし，Fig.6 左側に示すように，各センサ空間での直線経路

が必ずしも効率のよい経路であるという保証はない．例えば制御入力のノルムなどを計量として用いて最短経路を求めるなど，最適制御を導入することで経路の効率化を図ることが課題の一つとして考えられる．

5. おわりに

本研究では，観測範囲に制限のある複数のセンサ情報を用いた行動生成法を提案した．拡散学習にもとづき，観測範囲内で得られた Jacobi 行列を外延することにより仮想的なセンサ変数空間を観測範囲外に構築することを提案した．これにより，複数の（観測範囲の重ならない）センサ変数空間同士の対応関係を記述し，片方のセンサ変数空間から別のセンサ変数空間への動作生成を実現した．提案手法をマニピュレータによる床へのリーチング行動と移動ロボットによる壁へのナビゲーションという2種類の例題において検証を行った．

提案手法をセンサ情報に冗長性のある場合を含めた，より高次元の状態空間の問題に適用することも可能であり，それは今後の研究課題である．センサ信号のうちどれを選択することで適切な状態表現ができるのかという問題を考えることで，より多様で複雑なセンサ配置をもったロボットにとっての自律的な行動生成が可能になると考えられる．

謝辞

本研究の一部は財団法人テレコム先端技術研究支援センター研究費助成の支援を受けて行われた．ここに謝意を表します．

参考文献

- [Asada 01] Asada, M., MacDorman, K., Ishiguro, H., and Kuniyoshi, Y., Cognitive developmental robotics as a new paradigm for the design of humanoid robots. Robotics and Autonomous Systems, Vol. 37, 185-193(2001)
- [Luo 02] R. C. Luo, C. C. Yih, and K. L. Su, "Multisensor fusion and integration: approaches, applications, and future research directions," IEEE Sensors Journal, 2-2, 107/119(2002)
- [Thrun 05] S. Thrun, W. Burgard, and D. Fox, Probabilistic robotics, 2005.
- [Kaelbling 98] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, "Planning and acting in partially observable stochastic domains," Artificial Intelligence, Vol. 101, 99/134 (1998)
- [McCallum 95] R. Andrew McCallum, "Instance-based utile distinctions for reinforcement learning with hidden state," In Proceedings of the 12th International Machine Learning Conference, 1995.
- [木村 96] 木村 元, 山村 雅幸, 小林 重信, "部分観測マルコフ決定過程下での強化学習: 確率的傾斜法による接近," 人工知能学会誌, 11-5, 761/768 (1996)
- [Luo 98] Z. W. Luo and M. Ito, Diffusion-based learning theory for organizing visuo-motor coordination, Biological Cybernetics, Vol. 79, 279-289 (1998)