NAIST-IS-MT1351087

修士論文

遠隔操作型ロボットハンドのための
触覚情報に基づく器用さ支援方策の強化学習

長谷川　嵩大

2015 年 3 月 12 日

奈良先端科学技術大学院大学
情報科学研究科 情報科学専攻

本論文は奈良先端科学技術大学院大学情報科学研究科に
修士 (工学) 授与の要件として提出した修士論文である。

長谷川　嵩大

審査委員：

杉本　謙二	教授	（主指導教員）
小笠原　司	教授	（副指導教員）
松原　崇充	助教	（副指導教員）
南　裕樹	助教	（副指導教員）

遠隔操作型ロボットハンドのための触覚情報に基づく器用さ支援方策の強化学習 *

長谷川　嵩大

内容梗概

オペレータがロボットハンドを操作して特定のタスクを実行する際，タスクの達成のために，操作するロボットハンド周辺の環境情報を必要とする．この環境情報が欠如していた場合，タスク達成に必要な動作に過不足が発生し，タスクが達成できない可能性がある．この問題に対して，オペレータからの指令値を，環境情報を用いて補正する Shared Control という手法を用いた研究がなされている．しかしながら，従来の研究では，取り扱う問題や環境のモデル化を行い，パラメータを解析的または実験的に決定する必要があり，モデル化誤差が課題であった．そこで本研究では，強化学習の枠組みを用いてパラメータを自律的に設計する枠組みを提案する．提案手法の有効性の確認のために，ページめくりタスクを再現した物理シミュレーションを作成し，方策の学習とオペレータによるページめくり実験を行った．学習の結果，期待報酬が上昇する様なパラメータが学習された．このパラメータを用いて，オペレータによるページめくり実験を行った結果，ページめくりの支援によるタスク達成度の上昇が確認された．オペレータがシステムに含まれる Shared Control の利点を生かして，多様なページめくり動作についても学習された方策が機能することを確認した．

キーワード

ヒューマンロボットインタラクション, シェアードコントロール, 触覚情報, 強化学習, EM 方策探索法

*奈良先端科学技術大学院大学 情報科学研究科 情報科学専攻 修士論文, NAIST-IS-MT1351087, 2015 年 3 月 12 日.

Reinforcement Learning of Dexterity Assist for Tele-operated Robotic Hand Based on Tactile Information*

Takahiro Hasegawa

Abstract

The environment information obtained through several sensors is important for a tele-operated robotic hand to accomplish the dexterous manipulation task. However, the human operator cannot obtain all of the environmental information due to remote operation. If significant information is missing, it may result in too much or little control input to the robot. Therefore, the robot cannot achieve the task properly. Regarding this issue, one promising approach called shared control has been studied, which combines some of the autonomy with sensory information to the master-slave bilateral system with a human operator. However, previous studies are regarded as model-based methods, i.e., they require the complete knowledge of the task and environment, that may be unrealistic assumption and there are several concerns such as modeling errors. Therefore, in this paper, we propose a novel model-free method for designing shared control using reinforcement learning through trial and errors. To validate our method, we applied it for a page-turning task with a physical simulator of a paper sheet and a point-mass-like simplified robot fingertip. As a result of learning the significant improvement in terms of the accumulated reward is observed. Moreover, the subjective experimental results suggest that the learned policy can assist the

*Master's Thesis, Department of Information Science, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT1351087, March 12, 2015.

dexterity of a variety of page turning skills by taking the advantage of shared control with a human operator.

Keywords:

Human-Robot Interaction, Shared Control, Tactile Information, Reinforcement Learning, EM-Policy Search

目 次

図 目 次

表 目 次

1. 序論

1.1 はじめに

現在，オペレータが遠隔操作型のロボットを用いてより幅広いタスクを効率よく実行するために，遠隔操作型ロボットハンドの器用さの向上に関する研究が盛んに行われている．本研究では，オペレータによるこのロボットハンドの操作を支援する枠組みについて扱う．

1.2 環境情報と遠隔操作型ロボットハンドの器用さ

従来より，精密な動作を可能にする指先の器用な遠隔操作型ロボットハンドの研究・開発が行われてきた．これらの技術が活用される機器の例として，前腕切断者の Quality of Life(QOL) を向上させるための福祉機器である筋電義手等が挙げられる．しかし，オペレータによる遠隔操作型ロボットハンドの操作には，遠隔地にあるロボットハンドの置かれている環境をオペレータが十分に観測できないことに起因する課題が存在する．例えば，オペレータが遠隔操作型ロボットハンドを操作して物体把持タスクを行う場合，力の加えすぎによる物体もしくはロボットハンドの破損や，力の不足によるタスクの失敗がこれに当てはまる．この場合，ロボットハンドの触覚情報という環境情報の欠如が，タスクの達成を妨げている．この問題に対して，Shared Control という制御手法を用いて，遠隔操作型ロボットハンドの器用さを向上させ，与えられたタスクの成功率を上昇させる研究が盛んに行われている．Shared Control とは，オペレータと機械が協調的にロボットの制御を行う手法である．これによって，オペレータと機械が，それぞれ持っている利点を活かした制御が可能となる．

関連研究は次のようなものが存在する．Sherstan らは，オペレータの発生させる入力からロボットハンドがオペレータの意思を予測し，その予測に基づいて遠隔操作型ロボットハンドの制御を行うという枠組みを提案し，実機実験を通してその有効性を確認した [1]．Dragan らは，オペレータの発生させる入力からロボットアームがオペレータの意思を予測し，ロボットアームによる物体把持タス

クの達成を支援する枠組みを提案し，実機実験を通してその有効性を確認した [2]. Marayoug らは，オペレータと機械の協調的な動作の生成による物体把持タスクの枠組みを提案し，物理シミュレーションを通してその有効性を確認した [3].

しかしながら，これらの関連研究では方策のパラメータを環境や与えられたタスクのモデル化に基づいて決定しているため，柔軟物体などモデル化が困難な対象への適用は難しい．そこで，本研究では，Shared Control の枠組みの応用可能範囲の拡大を目的として，環境情報を利用した遠隔操作型ロボットハンドの器用さを向上する方策を強化学習を用いて自律的に設計する新しいアプローチについて検討する．

1.3 研究目的

本研究では，Shared Control の思想に基づいて，オペレータによる遠隔操作型ロボットハンドの操作を支援し，遠隔操作型ロボットハンドの器用さを向上する方策を自律的に獲得する枠組みの提案を目的とする．本研究において，ロボットハンドの器用さとはロボットハンドの動作の精度のことを指し，器用さの不足とは，与えられたタスクに対するロボットハンドの動作の過不足を指す．器用さの支援とは，この動作の過不足分を制御器が補正する事を意味する．この器用さ支援方策の構築に対して，環境や高次元ロボットハンド等の複雑な対象のモデル化を行わず，遠隔操作型ロボットハンドが環境との相互作用を通じて自律的に方策を設計するという，従来技術とは異なる思想に基づくアプローチを検討する．この様な方策を獲得する手法として，強化学習を用いる．提案手法の有効性の確認は，物理シミュレーションにて行う．

1.4 本論文の構成

本論文の構成について述べる．

2 章では器用さ支援について説明し，3 章で提案する Shared Control の枠組みについて述べる．4 章で本研究における提案法の適用について述べ，5 章では提案法

の有効性を確認するための実験とその結果について述べる．6 章では提案法についての考察と今後の展望について述べる．

2. Shared Controlに基づく支援方策

本章では，Shared Controlの概要と，本研究で提案する遠隔操作型ロボットハンドの器用さ支援方策の着想について述べる．本研究で扱う方策の構成図を，図1に示す．図1中の方策 (Policy) は，システムの状態が与えられた際に，システムへの制御入力を決定するものを表す．図1中のシステム (System) は，遠隔操作型ロボットハンドと環境の両方を含んだものを指す．与えられたタスクを達成できるように方策のパラメータを設定することが，器用さ支援方策を作成する上での目的となる．

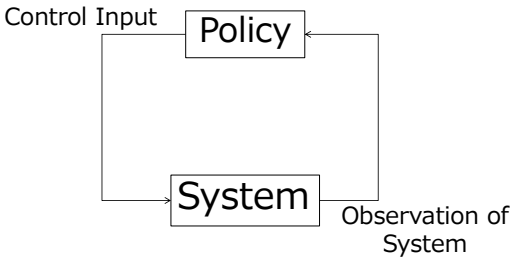


図 1　ブロック線図：System

この構成をShared Controlの枠組みで表すと図2の様になる．図2から，オペレータとコントローラの両方が方策に含まれることがわかる．なお，オペレータの方策のみを扱うManual Control，コントローラの方策のみを扱うAutomatic Controlと呼び，Shared Controlはこれらの手法を合わせたものであるといえる．

また，オペレータの方策とコントローラの方策はそれぞれ異なる特徴を持つ．オペレータの方策の特徴をManual Controlの解説と，コントローラの方策の特徴をAutomatic Controlの解説と共に示す．

2.1 Manual Control

Manual Controlは，オペレータが観測したシステムの状態に応じて入力を生成し，その入力がそのまま遠隔地のロボットハンドに反映される制御手法である．

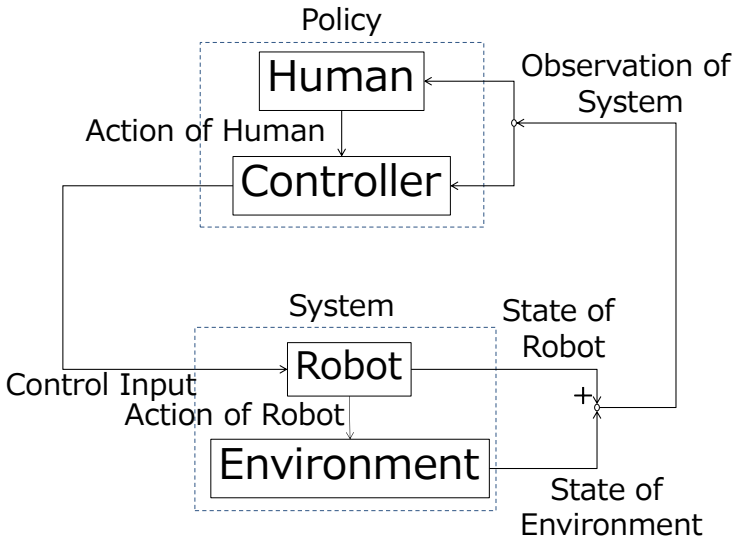


図 2 ブロック線図：Shared Control

この手法のブロック線図を図 3 に示す．この枠組みでは，オペレータがロボットハンドの動作を決定する方策となる．

この特徴によって，オペレータの入力と遠隔地のロボットハンドの動作の対応関係の把握が容易となり，オペレータは直観的にロボットハンドを操作することが可能となる．また，オペレータの持つ高次の認識を積極的に利用することで，複雑なタスクに対する動作計画が容易であるという利点が存在する．

一方で，この手法には，オペレータが遠隔地のロボットハンドの置かれている環境情報を十分に観測できない場合，ロボットハンドの器用さが大きく損なわれるという課題が存在する．

これらの課題に対して関連研究では，センサを増設することでオペレータが活用できる情報を増加させ，遠隔操作型ロボットハンドの操作精度の向上を図る研究がなされている．Vogel らは，オペレータの腕のトラッキングを行い，それをロボットハンドへの入力として扱うことによってオペレータから入力される情報を増やし，遠隔操作型ロボットハンドの操作精度の向上を実現した [4]．Yokoi らは，ロボットハンドの指先に触覚センサを設置し，その情報を電気信号に変換し

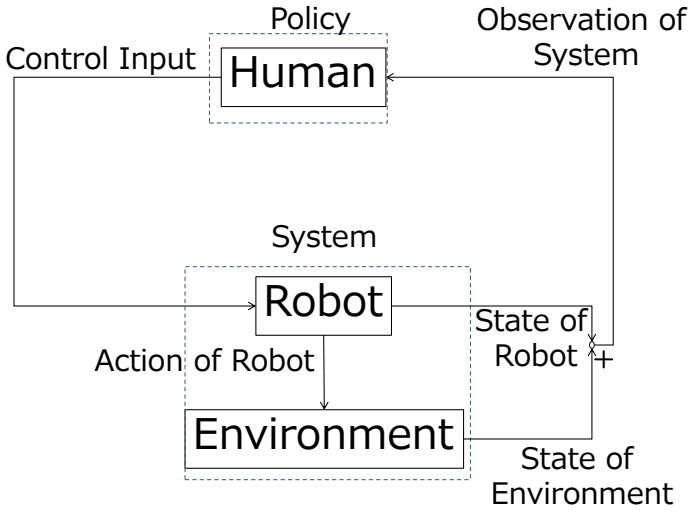


図 3　ブロック線図：Manual Control

てオペレータに呈示することで遠隔操作型ロボットハンドによる物体把持タスクの精度向上を実現した [5]．Tachi らは，ロボットの触覚情報をオペレータにフィードバックする装置を開発し，遠隔操作型ロボットの作業精度の向上を実現している [6]．Rombokas らは，シミュレーション上のロボットハンドの触覚情報をオペレータにフィードバックすることで，仮想的な物体の操作タスクの精度向上を実現している [7]．

2.2 Automatic Control

Automatic Control は，センサ情報を基に，コントローラが入力を決定する制御手法である．この手法のブロック線図を図 4 に示す．この枠組みでは，コントローラがロボットハンドの動作を自律的に決定する方策となる．

この特徴によって，センサによってオペレータ以上に正確に環境情報を利用できるため，オペレータ以上に正確にタスクの達成ができるという利点が存在する．また，コントローラによって高速で情報処理が行われるため，オペレータよりも素早くタスクの達成ができるという利点が存在する．

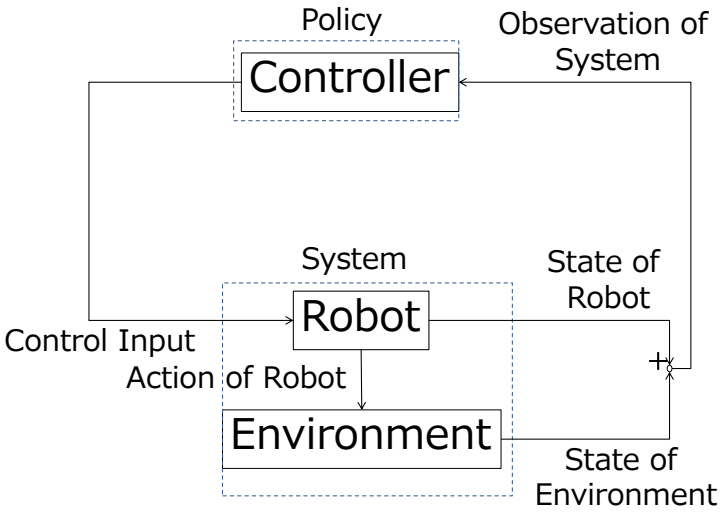


図 4　ブロック線図：Automatic Control

一方で，この手法では，環境のモデル化が難しい場合に方策のパラメータ調整が困難になるという課題が存在する．また，環境のモデル化が可能であったとしても，事前に設定されている動作以外を生成できず，動作の柔軟性に欠けるという課題が存在する．

これらの課題に対して，強化学習を用いてパラメータの調整を行う研究や，動作計画の充実による動作の柔軟性を増す研究がなされている．Sauser らは，触覚情報に基づいて，与えられた物体の把持に必要な握力を，自律型ロボットハンドが学習する枠組みを提案した．この枠組みを用いて，実際の自律型ロボットハンドによる適切な握力の呈示を実現した [8]．Ueda らは，自律型ロボットハンドの指先に触覚センサを設置し，その情報を基にロボットハンドが最適な動作を学習する枠組みを提案した．この枠組みを用いて，実際の自律型ロボットハンドによるページめくりタスクを実現した [9]．Daoud らは，触覚センサを装着した自律型多指ロボットハンドによる素早い把持動作の動作計画法 [10] や，実時間での把持動作の動作計画法 [11] を提案し，実機を用いてその有効性を確認した．Saut らは，複雑な物体に対する物体把持を行う際の自律型ロボットハンドの動作計画の

枠組みを提案した．この枠組みを用いて，複雑なモデルを持つ物体把持の実現を行った [12].

2.3 Shared Control の特徴と課題

2.1 節，2.2 節で述べたように，オペレータの方策では高次元の認識が，センサを用いた方策では詳細な環境情報が活用できる．すなわち，これらを組み合わせた手法である Shared Control には，オペレータの持つ高次の認識と機械の持つ情報処理能力を組み合わせることができるため，複雑な動作やタスクを素早くかつ正確に達成できるというを利点が存在する．

一方で，この手法では方策への入力が高次元になるため，方策パラメータの決定が困難であるという課題と，オペレータの入力が不確実性を含むため，オペレータの意思と入力の間の誤差が増大しやすいという課題が存在する．

これらの課題に対して，特定のタスクを想定して方策パラメータの決定を行った研究や，オペレータの入力を有限オートマトンによって抽象化し，オペレータの意思との誤差の影響を抑制する研究がなされている．Kroemer らは，オペレータの入力に対してカメラからの情報に基づいた動作の補正を行い，把持する物体に対して適切な動作を計画する枠組みの提案を行った．この枠組みを実機に適用し，6 種類の形状を持つ物体に対する物体把持タスクの精度向上を実現した [13]. Griffin らは，オペレータによる遠隔操作型ロボットハンドの触覚情報と視覚情報の観測に基づくオペレータの動作と機械的なコントローラによるロボットハンドの触覚情報の観測に基づいてロボットハンドの動作を決定する制御器を提案し，遠隔操作型ロボットハンドの物体把持タスクにおける器用さ向上を実現した [14]. Christian らは，オペレータの入力である筋電位信号 (EMG) に対してオートマトンを用いたオペレータの意思の離散的な推定を行い，オペレータの望むロボットハンドの状態を正確に決定する手法を提案した．これによって，Christian らは EMG ロボットハンドによる物体把持タスクの成功率向上を実現した [15].

2.4 提案するシステムの着想

遠隔操作型ロボットハンドの器用さの向上を実現するために必要となるのは，環境情報の詳細な観測と柔軟な動作計画である．この二つの課題に対して，オペレータによる対応と，センサを用いた場合による対応を比較する．

環境情報の観測は，センサを用いた方が詳細な環境情報を得られる．これは，センサによる環境の計測と比べて，遠隔操作を行っているオペレータには観測できる情報が少ないことに起因する．すなわち，環境を観測する点においては，オペレータではなくセンサが行った方が有効である．

一方，動作計画においては，オペレータの動作計画の方が柔軟性が高い．これは，オペレータの持つ高次の認識を用いた動作計画と異なり，コントローラによる制御は，あらかじめ設定された規範に従って行動することに起因する．すなわち，動作計画を行う点においては，ロボットではなくオペレータが行った方が有効である．

以上のことから，本研究では，オペレータによる動作計画とセンサによる環境の観測に基づいて協調的にロボットハンドへの入力を決定する枠組みの開発を行う．しかしながら，一般に，そのような方策を設計することは困難である．その理由として，環境の正確なモデル化や，高精細な動作を表現する規範の設計が困難なためである．

そこで提案手法では，強化学習の枠組みを用いて，ロボットハンドがオペレータと環境との相互作用を通じて自律的に学習する方針をとる．強化学習によって，オペレータの行動方策も考慮し，かつ，環境のモデル化を一切行わずに，行動の結果与えられる報酬が最大になる最適な動作を生成する方策を獲得する．

本研究では，強化学習手法に EM 方策探索法を用いる．EM アルゴリズムは，確率モデルのパラメータを最尤法に基づいて推定する手法のひとつで，期待値を計算する E ステップとその最大化を行う M ステップを交互に繰り替えすことで解を求めるものである．これによって，報酬関数が複雑で，直接最大化できない場合であっても効率的に解を求めることが可能となる．

3. 提案法：強化学習による Shared Control の学習法

本章では，提案法である強化学習による Shared Control の学習法について述べる．

3.1 提案法

2.3 節で述べたように，Shared Control の枠組みでは方策パラメータの決定が困難であるという課題が存在する．そこで，本研究では，ロボットハンドが環境との相互作用を通じて自律的に Shared Control の方策パラメータを学習する枠組みを提案する．

提案法のブロック線図を図 5 に示す．オペレータの方策にはオペレータの感覚を通した環境の観測が，支援方策にはオペレータの方策が出力したオペレータによる制御入力とセンサによる環境の観測が，それぞれ入力される．これによって，オペレータの高次の認識とセンサによる詳細な環境の観測を利用することができる．この構成では，オペレータの観測に基づいてオペレータの方策が生成した制御入力に対して，センサの観測に基づいて支援方策が補正を行う枠組みとなっている．また，方策を用いた時の入出力関係，試行を行った時に得られた報酬の記録する．一定回数の試行毎に，強化学習の枠組みで支援方策を更新し，最適なパラメータを学習する．

3.2 変数の設定

$\mathcal{S}$ は状態集合を，$\mathcal{A}$ は行動集合を，$\mathcal{O}$ は観測集合を，$P_T(s'|s,a)$ は状態遷移モデルを，$P_o(o|s)$ は観測モデルをそれぞれ表す．さらに，$\pi(a|o;\theta)(\leq 0)$ はパラメータ θ を持つエージェントの方策モデルである．

状態 $\mathcal{S} \ni \boldsymbol{s}$ は，ロボットハンドの状態と環境の状態を含む．ロボットハンドの行動 $\mathcal{A}^c \ni \boldsymbol{a}^c$ は，支援方策 π^c によって決定される．オペレータの行動 $\mathcal{A}^h \ni \boldsymbol{a}^h$ は，オペレータの行動方策 π^h によって決定される．センサによる環境の観測 $\mathcal{O}^s \ni \boldsymbol{o}^s$

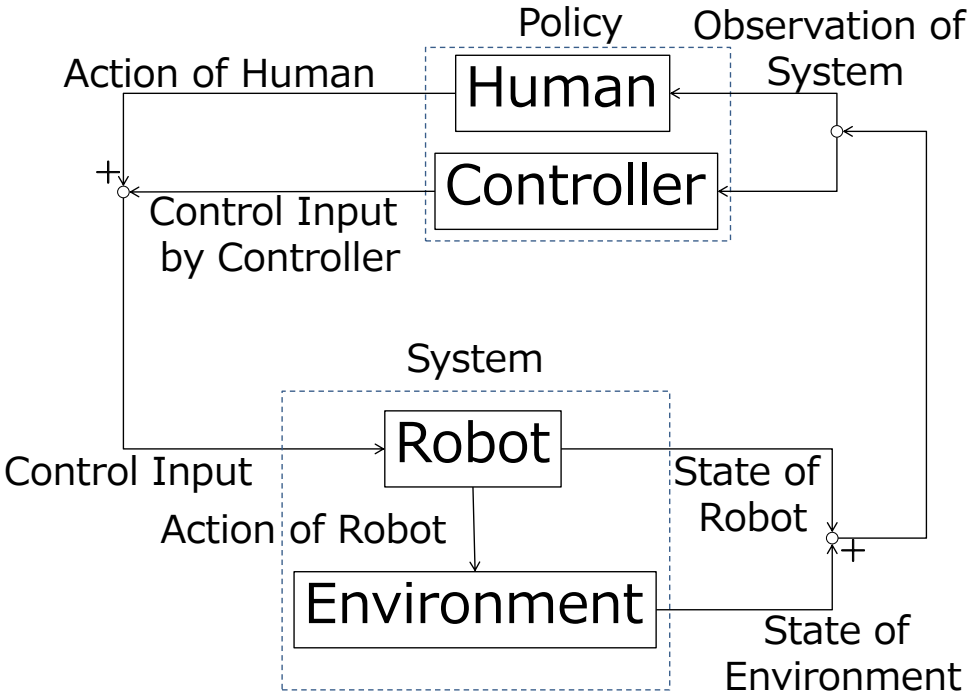


図 5 ブロック線図：提案法

は，ロボットハンドに装着された触覚センサによって決定される．オペレータによる環境の観測 $\mathcal{O}^s \ni \boldsymbol{o}^s$ は，オペレータの視覚によって決定される．

これらの変数 $(\boldsymbol{s}, \boldsymbol{a}^h, \boldsymbol{a}^c, \boldsymbol{o}^s)$ の変化をエピソード $\boldsymbol{d}$ とする．

3.3 器用さ支援方策の設定

本節では，器用さ支援システムの設定について述べる．2.4 節で述べた，オペレータによる遠隔操作型ロボットハンドへの制御入力の過不足をシステムが補正する枠組みの詳細について述べる．ロボットハンドへの制御入力は，オペレータの行動と，支援行動の和によって決定されるとする．

このロボットハンドへの制御入力 $\boldsymbol{a}^r$ を決定するオペレータの行動方策と，支援方策のモデルについて解説する．

まず，オペレータの行動 $\boldsymbol{a}^h$ を決定するオペレータの行動方策 π^h について述べる．オペレータの行動方策 π^h は，オペレータによる環境の観測 $\boldsymbol{o}^h$ を入力として用いて，式 (1) の様な確率モデルで表される．

$$\pi^h(\boldsymbol{a}^h|\boldsymbol{o}^h) \tag{1}$$

次に，支援行動 $\boldsymbol{a}^c$ を決定する支援方策 π^c について述べる．支援方策 π^c は，ロボットハンドの状態 $\boldsymbol{s}^r$，オペレータの行動 $\boldsymbol{a}^h$，センサによる環境の観測 $\boldsymbol{o}^s$ を入力として用いる．また，支援行動 $\boldsymbol{a}^c$ は，入力の他に方策パラメータ $\boldsymbol{\theta} = (\boldsymbol{w}, \boldsymbol{\sigma})^T$ によって決定される．以上のことから，支援方策 π^c は，式 (2) の様な確率モデルで表される．

$$\pi^c(\boldsymbol{a}^c|\boldsymbol{s}^r, \boldsymbol{a}^h, \boldsymbol{o}^s; \boldsymbol{\theta}) \tag{2}$$

以降では，$(\boldsymbol{s}^r, \boldsymbol{a}^h, \boldsymbol{o}^s)$ を $\boldsymbol{x}$ と略記する．本研究では支援方策 π^c に，次のガウシアン方策モデルを仮定する．

$$\pi^c(\boldsymbol{a}^c|\boldsymbol{x}; \boldsymbol{\theta}) \;=\; \frac{1}{\sqrt{2\pi\boldsymbol{\sigma}^2}} \exp\left(-\frac{(\boldsymbol{a}^c - \boldsymbol{w}^{\mathrm{T}}\boldsymbol{\phi}(\boldsymbol{x})^2}{2\boldsymbol{\sigma}^2} \right) \tag{3}$$

式 (3) 中の $\boldsymbol{\phi} = (\phi_1(\boldsymbol{x}), \phi_2(\boldsymbol{x}), \cdots, \phi_B(\boldsymbol{x}))^T$ は基底集合，B は基底の数，方策パラメータは，$\boldsymbol{\theta} = (\boldsymbol{w}, \boldsymbol{\sigma})^T$ で構成される．

この枠組みにおいて，ロボットハンドへの制御入力 $\boldsymbol{a}^r$ は式 (4) で与えられる．支援行動はガウス基底関数の線形重み付け和によって与えられる．式 (4) 中の $\boldsymbol{\epsilon}$ は，平均 0 標準偏差 $\boldsymbol{\sigma}$ のガウスノイズを表す．

$$\begin{aligned} \boldsymbol{a}^r &= \boldsymbol{a}^h + \boldsymbol{a}^c \\ &= \boldsymbol{a}^h + \boldsymbol{w}^{\mathrm{T}}\boldsymbol{\phi}(\boldsymbol{x}) + \boldsymbol{\epsilon} \end{aligned} \tag{4}$$

このような方策モデルを用いた場合，全観測を仮定する [16] と同様に，EM 方策探索における M-step の計算は，報酬重み付き回帰問題に帰着できる．これにより，最適な方策パラメータ $\boldsymbol{\theta} = (\boldsymbol{w}, \boldsymbol{\sigma})^T$ を獲得する．

3.4 報酬関数

遠隔操作型ロボットハンドの Rollout $\boldsymbol{d}$ に基づく報酬値を，式 (5) のように設定する．式 (5) 中の $\gamma(\in (0,1])$ は未来の報酬の割引率を表す．

$$\mathcal{R}(\boldsymbol{d}) = \sum_{n=1}^{N} \gamma R(\boldsymbol{s}_n^r, \boldsymbol{a}_n^c, \boldsymbol{s}_{n+1}^r) \tag{5}$$

この時の各時刻の即時報酬は，式 (6) の形となる．

$$R(\boldsymbol{s}_n^r, \boldsymbol{a}_n^c, \boldsymbol{s}_{n+1}^r) > 0 \tag{6}$$

3.5 EM 方策探索法

パラメータの更新が L 回行われた時を考える．EM アルゴリズムに従って期待報酬 $J(\boldsymbol{\theta})$ よりも $\boldsymbol{\theta}$ に関する最大化が容易な下界 $J_L(\boldsymbol{\theta})$ の計算と，その最大化による $\boldsymbol{\theta}_{L+1}$ への更新を収束まで交互に繰り返すことで，最適な $\boldsymbol{\theta}^*$ を効率よく求める．

$\boldsymbol{\theta}_L$ を現在の方策パラメータとして，$\mathcal{R}(\boldsymbol{d})$ が非負であることに注意すると，Jensen の不等式を使って，期待報酬の下界に関する式 (7) を導出できる．

$$\begin{aligned} \log J(\boldsymbol{\theta}) &\geq \int \frac{\mathcal{R}(\boldsymbol{d})P(\boldsymbol{d};\boldsymbol{\theta})}{J(\boldsymbol{\theta}_L)} \log \frac{P(\boldsymbol{d};\boldsymbol{\theta})}{P(\boldsymbol{d};\boldsymbol{\theta}_L)} \mathrm{d}\boldsymbol{d} + \log J(\boldsymbol{\theta}_L) \\ &= \log J_L(\boldsymbol{\theta}) \end{aligned} \tag{7}$$

この下界の計算が E-step に, $\boldsymbol{\theta}$ に関する最大化問題を解き $\boldsymbol{\theta}$ を更新する以下の式 (8) が M-step にそれぞれ対応する．

$$\boldsymbol{\theta}_{L+1} = \arg\max_{\boldsymbol{\theta}} J_L(\boldsymbol{\theta}) \tag{8}$$

また $\log J_L(\boldsymbol{\theta}_L) = \log J(\boldsymbol{\theta}_L)$ のため (下界 J_L が $\boldsymbol{\theta}_L$ において期待報酬と接するため)，各反復で期待報酬が減少しないことが保障される．

$$J(\boldsymbol{\theta}_{L+1}) \geq J(\boldsymbol{\theta}_L) \tag{9}$$

方策パラメータ $\boldsymbol{\theta} = (\boldsymbol{w}, \boldsymbol{\sigma})^T$ の更新式は，[17] から式 (10)，式 (11) として導出される．式 (10)，式 (11) 中の変数 $\boldsymbol{a}$ は，この式中において支援行動 $\boldsymbol{a}^c$ を意味する．

$$\begin{aligned}\boldsymbol{w}_{L+1} =& \left(\frac{1}{M} \sum_{m=1}^{M} \mathcal{R}\left(\boldsymbol{d}_m^L\right) \frac{1}{N} \sum_{n=1}^{N} \boldsymbol{\phi}(\boldsymbol{x}_{m,n}^L) \boldsymbol{\phi}(\boldsymbol{x}_{m,n}^L)^{\mathrm{T}} \right)^{-1} \\ & \left(\frac{1}{M} \sum_{m=1}^{M} \mathcal{R}\left(\boldsymbol{d}_m^L\right) \frac{1}{N} \sum_{n=1}^{N} \boldsymbol{a}_{m,n}^L \boldsymbol{\phi}(\boldsymbol{x}_{m,n}^L) \right)\end{aligned} \tag{10}$$

$$\begin{aligned}\boldsymbol{\sigma}_{L+1}^2 =& \left(\frac{1}{M} \sum_{m=1}^{M} \mathcal{R}(\boldsymbol{d}_m^L) \right)^{-1} \\ & \left(\frac{1}{M} \sum_{m=1}^{M} \mathcal{R}(\boldsymbol{d}_m^L) \frac{1}{N} \sum_{n=1}^{N} \left(\boldsymbol{a}_{m,n}^L - \boldsymbol{w}_{L+1}^{\mathrm{T}} \boldsymbol{\phi}(\boldsymbol{x}_{m,n}^L) \right)^2 \right)\end{aligned} \tag{11}$$

3.6 予備実験

ここでは，EM 方策探索法の有効性を確認する．本実験では，式 (12) のモデルを持つ DC モータの制御を行う．モータの角度，角速度を $\boldsymbol{x} = [\theta[\mathrm{rad}], \dot{\theta}[\mathrm{rad/sec}]]^{\mathrm{T}}$，入力電圧を u，モータの次のステップの状態を $\dot{\boldsymbol{x}}$ として設定する．

$$\dot{\boldsymbol{x}} = \begin{bmatrix} 1 & 0.0952 \\ 0 & 0.9048 \end{bmatrix} \boldsymbol{x} + \begin{bmatrix} 0.2419 \\ 4.7581 \end{bmatrix} u \tag{12}$$

DC モータの初期角度を π[rad]，初期角速度を 0[rad/sec] とし，目標角度を 0[rad] とした時の DC モータの挙動を確認する．基底関数の平均は，モータの角度，角速度それぞれの状態が取りえる範囲を，各次元ごとに 10 等分した時の値の組み合わせで与える．今回設定した基底関数を図 6 に示す．

この予備実験では，EM 方策探索法を用いて最適な方策パラメータ $\boldsymbol{\Delta} = (\boldsymbol{w}^T, \sigma)^T$ を学習する．入力電圧 u は方策 π によって，式 (13) の様に，状態 $\boldsymbol{x}$ を入力とする基底関数 $\phi(\boldsymbol{x})$ の線形重みづけ和によって得られる．

$$u = \boldsymbol{w}^{\mathrm{T}} \boldsymbol{\phi}(\boldsymbol{x}) \tag{13}$$

また，この予備実験における即時報酬 r は式 (14) で表される．式 (14) 中の割引率 γ は 0.98，I は単位行列とする．

$$r = \gamma \frac{1}{(\boldsymbol{x}^{\mathrm{T}} I \boldsymbol{x} + u^2)} \tag{14}$$

1 Iteration における試行回数は 200 を設定した．この予備実験では，EM 方策探索法による期待報酬の上昇の確認と学習されたパラメータ $\boldsymbol{w}$ を用いたモータの挙動の確認を行う．学習は 3 回実施した．

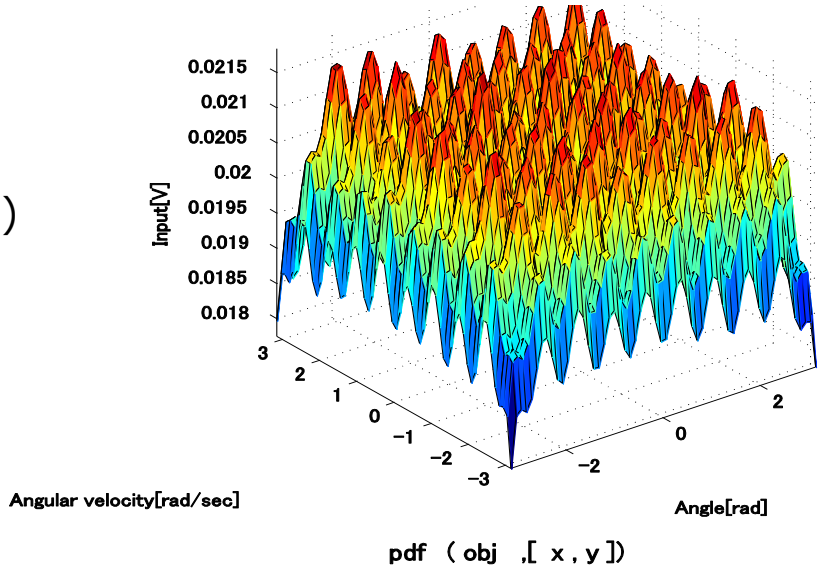
3.6.1 実験結果

Iteration 毎の期待報酬の推移を図 7 に示す．図 7 中の実線は 3 回の実験で得た期待報酬の平均を，エラーバーは 3 回の実験で得た期待報酬の標準偏差を表す．500 Iteration で期待報酬がおおよそ収束していることが確認できる．

500 Iteration 目に学習された方策を用いてモータを制御した際のモータの挙動を図 8 に示す．図 8 中 (a) は，モータの角度の推移を，(b) はモータの角速度の推移を，(c) は入力の推移を示す．図 8 から，角度，角速度が共に 0[rad]，0[rad/sec] に到達し，目標を達成していることが確認できる．

これによって，EM 方策探索による方策の学習が有効であることが確認された．

(a)


0.0215
0.021
0.0205
0.02
0.0195
0.019
0.0185
0.018
Input[V]
3
2
1
0
−1
−2
−3
Angular velocity[rad/sec]
−2
0
2
Angle[rad]

(b)

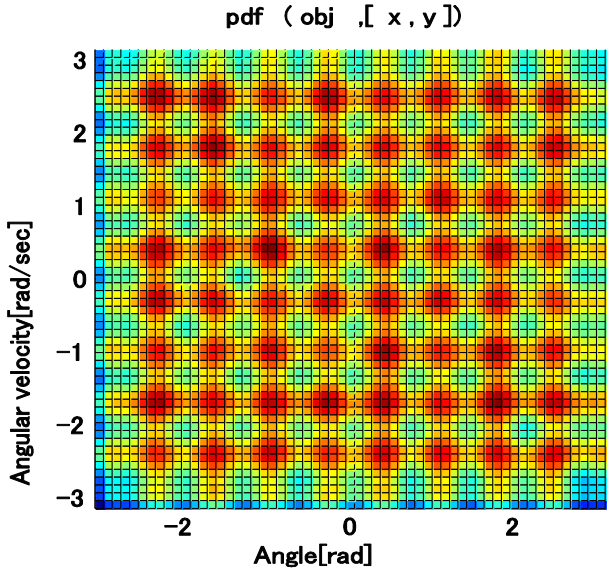

pdf （ obj ,[x , y]）
3
2
1
0
−1
−2
−3
Angular velocity[rad/sec]
−2
0
2
Angle[rad]

図 6 基底関数

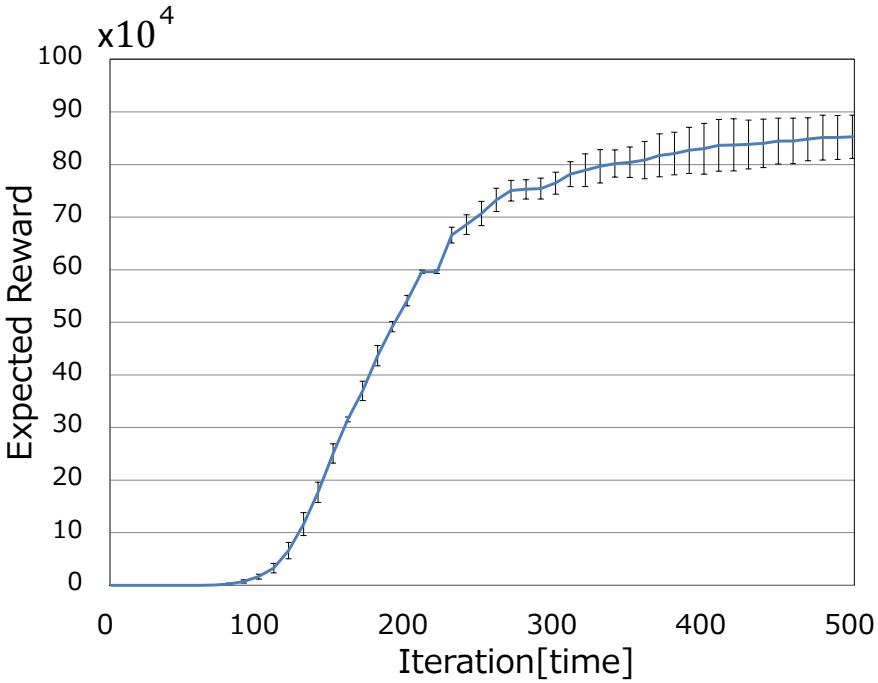

x10^4
100
90
80
70
60
50
40
30
20
10
0
Expected Reward
0
100
200
300
400
500
Iteration[time]

図 7 期待報酬

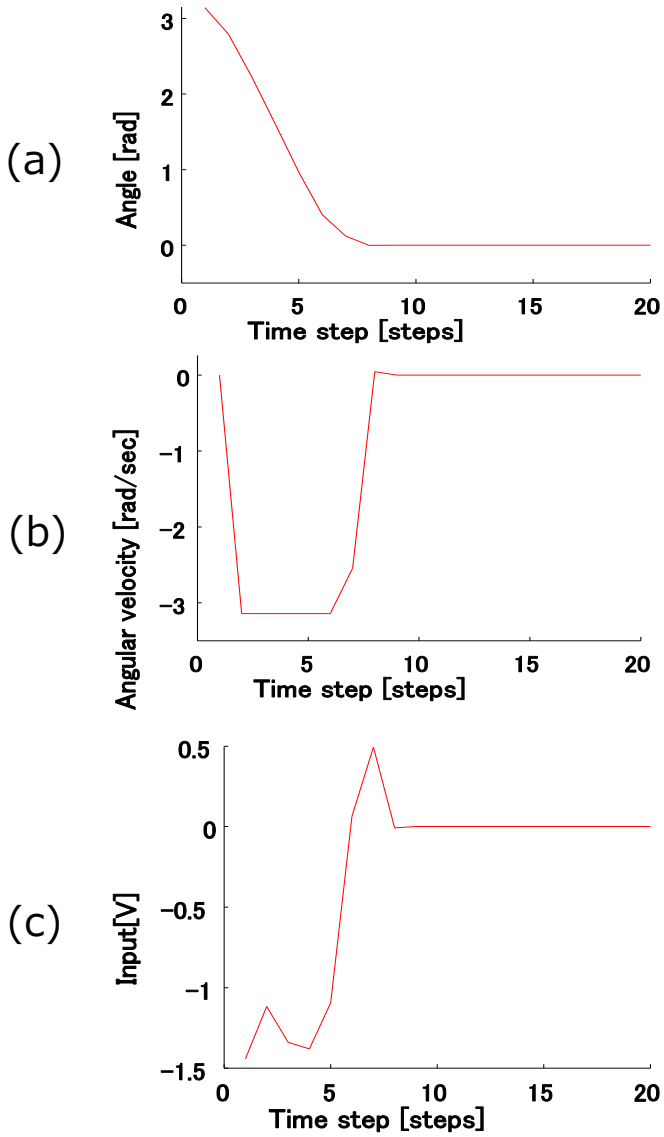

(a)
Angle [rad]
Time step [steps]
(b)
Angular velocity [rad/sec]
Time step [steps]
(c)
Input[V]
Time step [steps]

図 8　生成された挙動

4. ページタスクへの適用

3 章で示した提案手法を評価するための実験課題として遠隔操作型ロボットハンドを使ったページめくりタスク [9] を設定する．

4.1 ページタスク設定

本節では，器用さ支援方策を用いるページめくりタスクの設定について述べる．ここで，3.2 節で示した提案法の枠組みへの適用を行う．

状態として，ロボットハンドの状態 $\boldsymbol{s}^r$ と，ページの状態 $\boldsymbol{s}^p$ を考える．行動として，オペレータの行動方策 π^h によって決定される行動 $\boldsymbol{a}^h$ と，支援方策 π^c によって決定される行動 $\boldsymbol{a}^c$，ロボットハンドへの制御入力 $\boldsymbol{a}^r$ を考える．観測として，オペレータの行動観測 $\boldsymbol{o}^h$，オペレータによるページの状態観測 $\boldsymbol{o}^h_p$，オペレータによるロボットハンドの状態観測 $\boldsymbol{o}^h_r$， ページの状態観測 $\boldsymbol{o}^p$，触覚センサによる観測 $\boldsymbol{o}^{Tac}$ を考える．これらの変数の変化をエピソード $\boldsymbol{d}$ とする．

次に，オペレータと支援の行動方策 π^h，π^c の確率的なモデルについて述べる．オペレータの行動方策 π^h の入力である，オペレータによるロボットハンドの状態観測 $\boldsymbol{o}^h_r$ とオペレータによるページの観測 $\boldsymbol{o}^h_p$ は，ロボットハンドの状態 $\boldsymbol{s}^r$，ページの状態 $\boldsymbol{s}^p$ に基づいてそれぞれ式 (15)，式 (16) の様に確率的に変化するものとする．

$$p(\boldsymbol{o}^h_p|\boldsymbol{s}^p) \tag{15}$$

$$p(\boldsymbol{o}^h_r|\boldsymbol{s}^r) \tag{16}$$

確率の積の法則を用い，式 (17) が得られる．従って，ロボットハンドの状態 $\boldsymbol{s}^r$ とページの状態 $\boldsymbol{s}^p$ から，オペレータの行動 $\boldsymbol{a}^h$ を確率的に推定する枠組みを得る．

$$\pi^h(\boldsymbol{a}^h|\boldsymbol{o}^h_p, \boldsymbol{o}^h_r)p(\boldsymbol{o}^h_p|\boldsymbol{s}^p)p(\boldsymbol{o}^h_r|\boldsymbol{s}^r) \tag{17}$$

支援方策 π^a の入力である，オペレータの行動観測 $\boldsymbol{o}^h$ は，式 (18) の確率モデルに従う．

$$p(\boldsymbol{o}^h|\boldsymbol{a}^h) \tag{18}$$

従って支援方策の行動 $\boldsymbol{a}^c$ は式 (19) の確率モデルに従う．

$$\boldsymbol{a}^c = \pi^c(\boldsymbol{a}^c|\boldsymbol{s}^r, \boldsymbol{o}^h, \boldsymbol{o}^{Tac}; \boldsymbol{\theta})p(\boldsymbol{o}^h|\boldsymbol{a}^h)\pi^h(\boldsymbol{a}^h|\boldsymbol{o}^h_p, \boldsymbol{o}^h_r)p(\boldsymbol{o}^h_p|\boldsymbol{s}^p)p(\boldsymbol{o}^h_r|\boldsymbol{s}^r) \tag{19}$$

4.2 報酬関数

報酬関数を 3.4 節中の式 (5) に基づいて設定する．まず，各時刻の即時報酬は，式 (20) となる．

$$R(\boldsymbol{s}^r_n, \boldsymbol{a}^c_n, \boldsymbol{s}^r_{n+1}) \quad = \quad \lambda h - ||\boldsymbol{a}^c_n|| > 0 \tag{20}$$

この即時報酬は，与えられたのタスク達成度合いに基づく報酬と，支援に対する依存度合いに基づく罰則の 2 つで構成される．報酬は，本研究において，式 (20) 中の様に，ページのめくられた量 h を定数 λ によって定数倍することで与える．ページのめくられた量 h は，図 10 の様に持ち上げられたページの最も高い座標の高度の値である．

一方罰則は，本研究において，式 (20) 中の様に，各時刻における器用さ支援方策によるアシスト量 $\boldsymbol{a}^c$ のノルムによって与える．

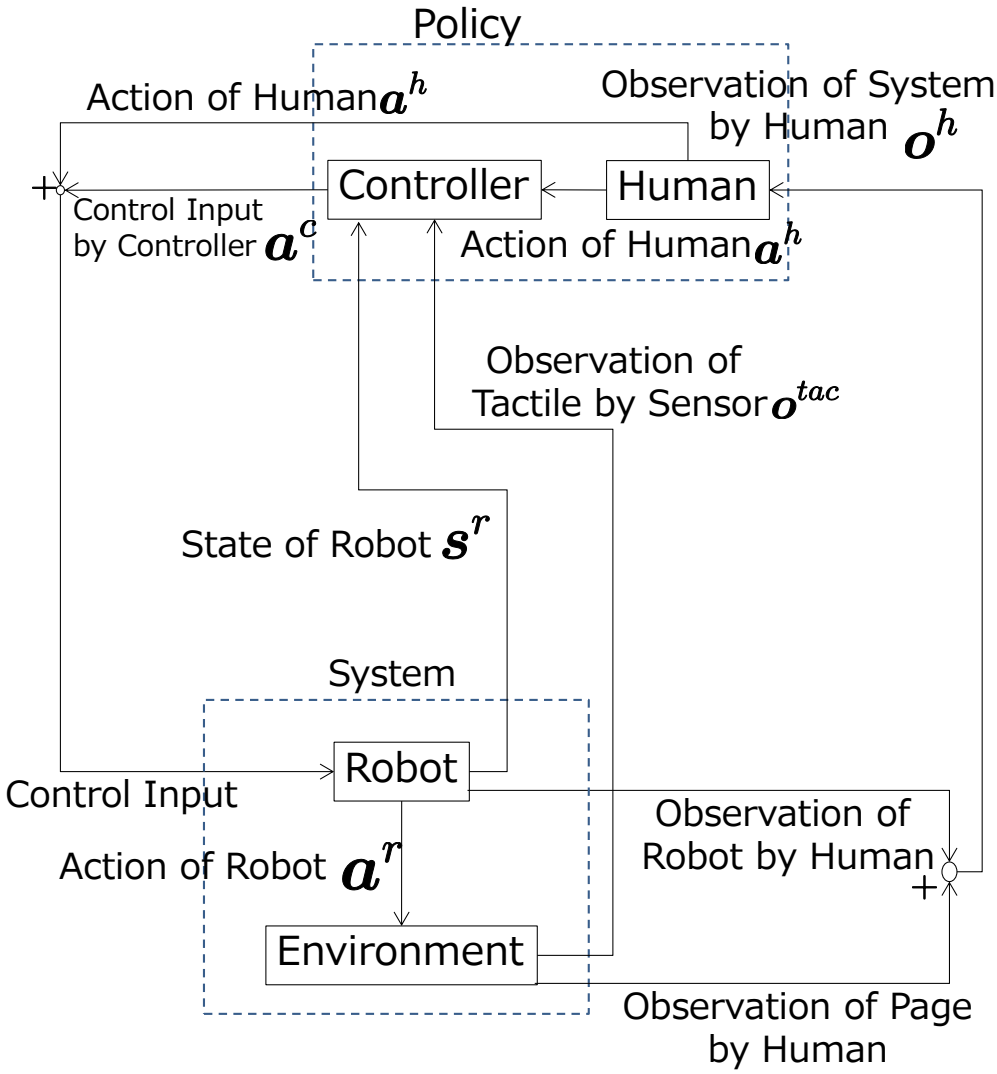


図 9　ブロック線図：ページめくりタスクに対する支援方策システム

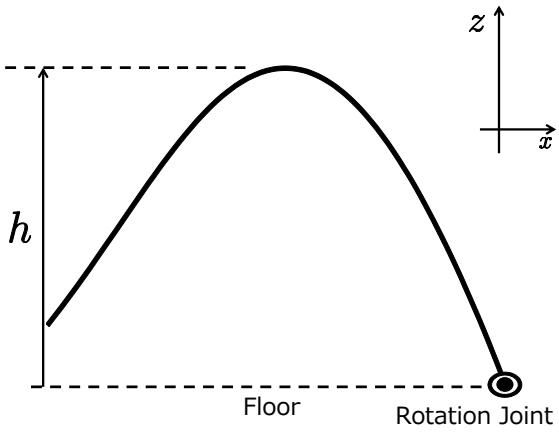


図 10　ページの高さの定義

5. 実験

本研究では，提案手法の有効性を，ロボットハンドによるページめくり実験を通して確認する．本章では，ページめくりタスクの実験環境，実験の設定，実験結果とそれに対する考察を行う．

5.1 実験環境

ここでは，多関節人工筋ロボットハンドによるページめくりタスクの実験環境について述べる．本研究では，多関節人工筋ロボットハンドとしてShadow Dextrous Hand(Shadow Robot 社) を対象とする (図 11)(以下，Shadow Hand に省略)．

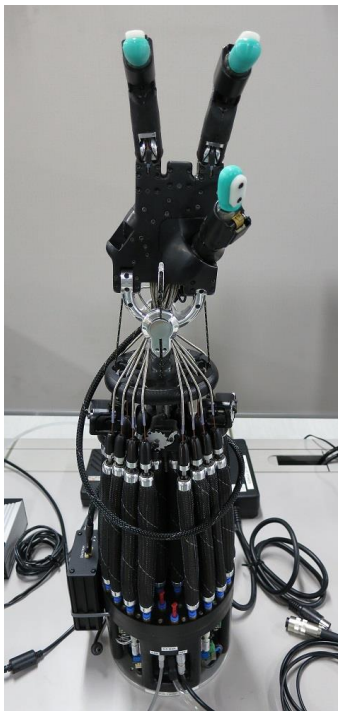


図 11 Shadow Hand

この Shadow Hand には，各指の先端に BioTac(SynTatch 社) という触覚センサが取り付けられ，圧力の検出を行う．圧力値はセンサ内部で 2200[Hz] の周期で

計測され，1040[Hz] のローパスフィルタを経由した圧力値を Pdc，10-1040[Hz] のバンドパスフィルタを経由した圧力値を圧力 Pac として取得することができる．各関節角度, 圧力は 10[ms] の取得周期で時系列の情報として保存される．Shadow Hand は高いコンプライアンス性を持ち，物体と接触した際，力がかかりすぎると他の関節角が柔軟に変化するため，学習中のロボットハンドの破損を抑制する効果が期待できる．Shadow Hand によるページめくりを行う実験環境を，図 12 に示す．

図 12　実験環境:実機

オペレータはページの状態と Shadow Hand の状態を観測しながら，Shadow Hand の操作を行う．この設定において，オペレータは Shadow Hand とページの接触情報を観測することができない．本研究では，学習のための時間の削減と，結果の再現性の確認のために，この設定を再現した物理シミュレーションを用いてコントローラの学習とページめくり実験を行う．

5.1.1 オペレータの行動観測に使用するデバイス

本研究では，オペレータの行動を観測するためのデバイスとして CyberGlove Systems 社製の cyberglove2(図 13) を利用した．このデバイスは，手に装着し，オペレータの指の関節角度や関節角速度を計測できるグローブ型のデバイスである．

図 13　入力デバイス:cyberglove

内蔵されているセンサは 1 本の指につき 3 つの曲げセンサー，4 つの外反センサー，手の平の反り（手掌弓），そして曲げと反りを計測するセンサが計 18 個あり，これによって，手や指の動きを関節角度のデジタルデータへリアルタイムで正確に変換する．本研究では主に親指の CM 関節の 2 自由度の計測を行い，Shadow Hand または物理シミュレーショへの入力とする．CM 関節は，親指の外転，内転，対立，復位の運動を行う際に稼働する関節である．図 14 に CM 関節の位置を示す．

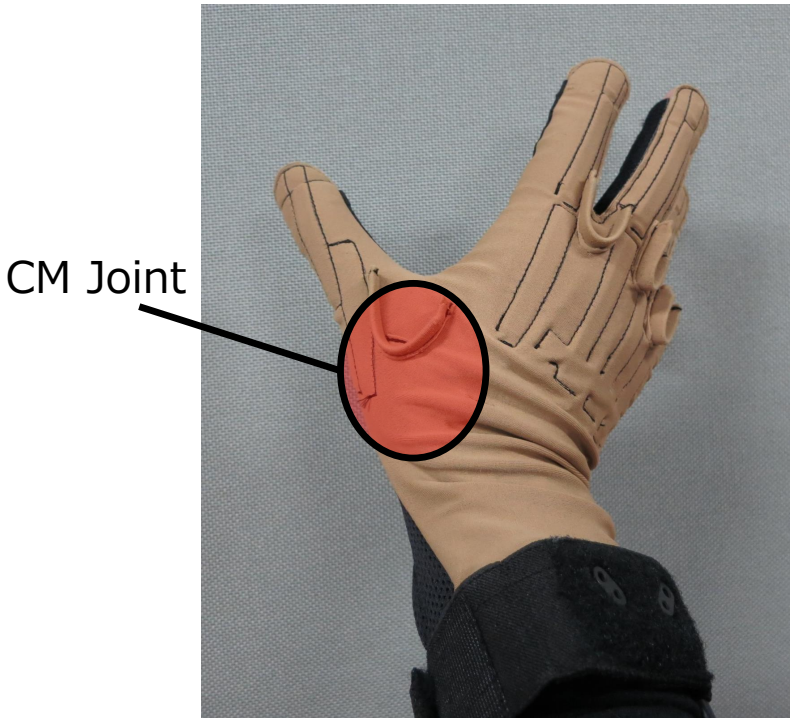


図 14　CM 関節

5.1.2 物理シミュレーション

ここでは，物理エンジン Open Dynamics Engine(ODE) を用いて作成した物理シミュレーションの解説を行う．シミュレータは，紙のページ，ロボットハンドの指先，シミュレータの状態を描画しオペレータの視覚に情報をフィードバックする 24 インチのディスプレイ，オペレータの指の関節角度を計測する cyberglove で構成される．これらのシステムは 1 台の PC(CPU:Intel Xeon CPU E5-1620(3.6GHz), Memory:8GB) の Robot-Operating-System(ROS) によって管理されている．物理シミュレーションの計算はシミュレータ内で 1kHz の周期で実施される．これは実環境における 100Hz と一致する．すなわち，シミュレータの計算周期は実環境よりも 10 倍遅いものである．また，シミュレータとオペレータのやり取りは 10Hz の周期で実施される．ODE を用いて作成したページのモデルを，図 15 に示す．

図 15 中の緑色の球体はロボットハンドの指先を表す．この球体は，画面中の縦

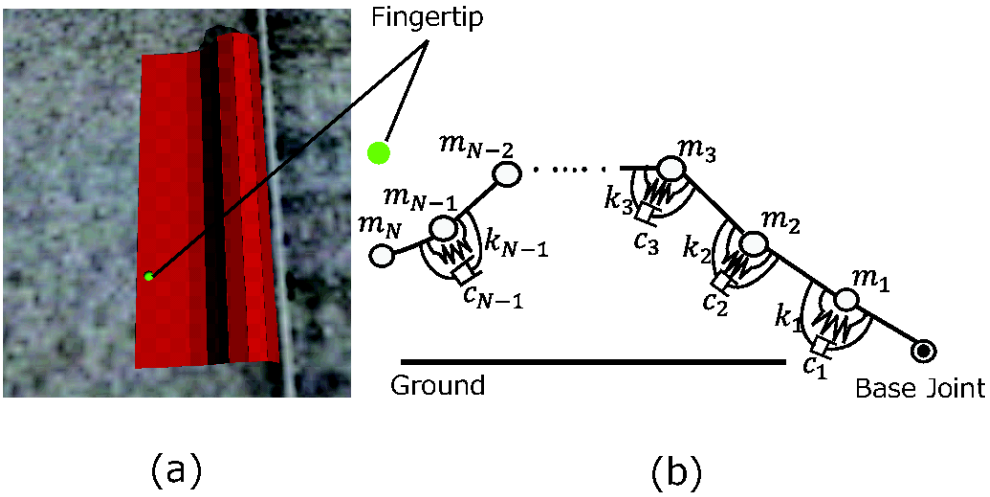


図 15　実験環境:物理シミュレーション

と横の 2 自由度をもつ．図 15 中の赤色の物体は，[9] に則り，ページを多リンクをもつアームによって近似したものである．図 15 にページのモデルを示す．図 15 中の変数 N は，ページのモデル作成に用いるアームの数である．本実験では，1[g]，1[cm] のアームを 10 個用いた．アームを接続するリンクは，画面に対して垂直の座標軸のみを回転軸として持つバネダンパ系として設定している．図 15 中の変数 k はバネ係数を，c はダンパ係数を表す．本研究では，実験的にバネ係数を 0.8，ダンパ係数を 0.001 に設定した．本実験では，オペレータからの入力を 2 つの関節角度に制限し，ロボットハンドの 2 自由度と対応付けを行い，直観的な操作を可能にした．図 16 に，オペレータの手の運動と物理シミュレータ内の指の運動の対応を示す．

5.1.3　実験環境の全体像

本実験の全体像を図 17 に示す．被験者は画面によって物理シミュレーション内のページの状態を確認し，自身の指の動作を生成する．この動作は cyberglove を用いて計測され，その値が被験者の入力として扱われる．この入力と物理シミュレーション内の指の状態，物理シミュレーション内の指の接触情報に応じて，支

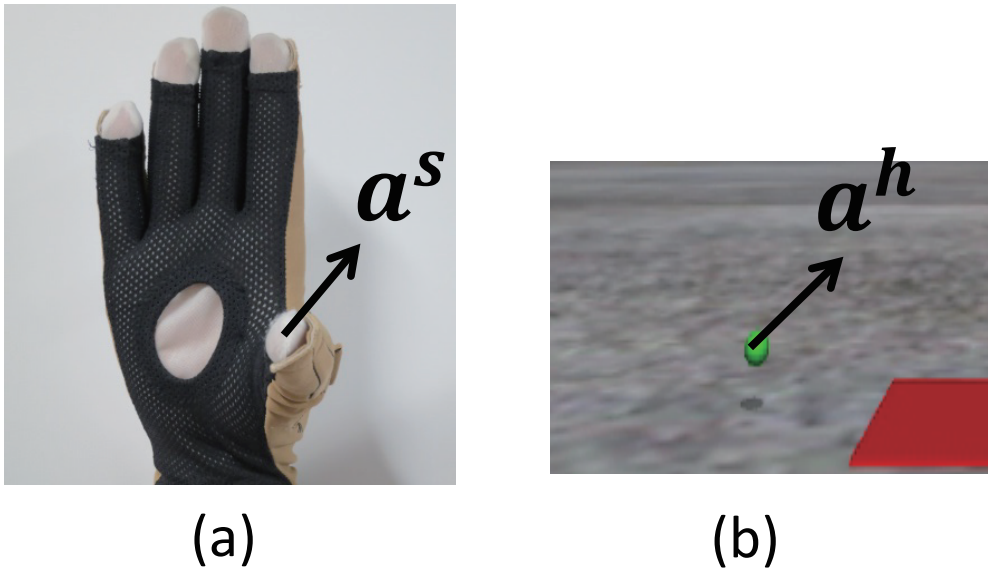


図 16　手の運動と物理シミュレータ内の指の運動の対応

援方策が補正量を決定する．被験者の入力と補正量の和が，物理シミュレーション内の指への入力となる．被験者は物理シミュレーション内の指を操作することによって，ページめくりタスクを実施する．

5.2　オペレータの行動方策モデル

本実験では，被験者の負担軽減のために，器用さ支援方策の学習に際して，被験者による入力を疑似的に再現するオペレータの行動方策モデルを作成した．この方策モデルは，現時刻での物理シミュレーション内の指の座標，時間，ページのめくり度合いから次のステップのオペレータの入力を確率的に決定するものである．本研究では，式 (21) の様に，基底関数の線形重み付けによってオペレータの行動を生成する，オペレータの行動方策モデルを設計する．式 (21) 中の変数の設定は次の通りである．基底関数の入力として，物理シミュレーション内の指の状態 $\boldsymbol{s}^r$，実験時間 t，ページの状態 $\boldsymbol{s}^p$ を， 出力となる行動変数 a として，$\boldsymbol{a}^h$ をそれぞれ設定する．基底関数には物理シミュレーション内の指先の 2 次元座標，経

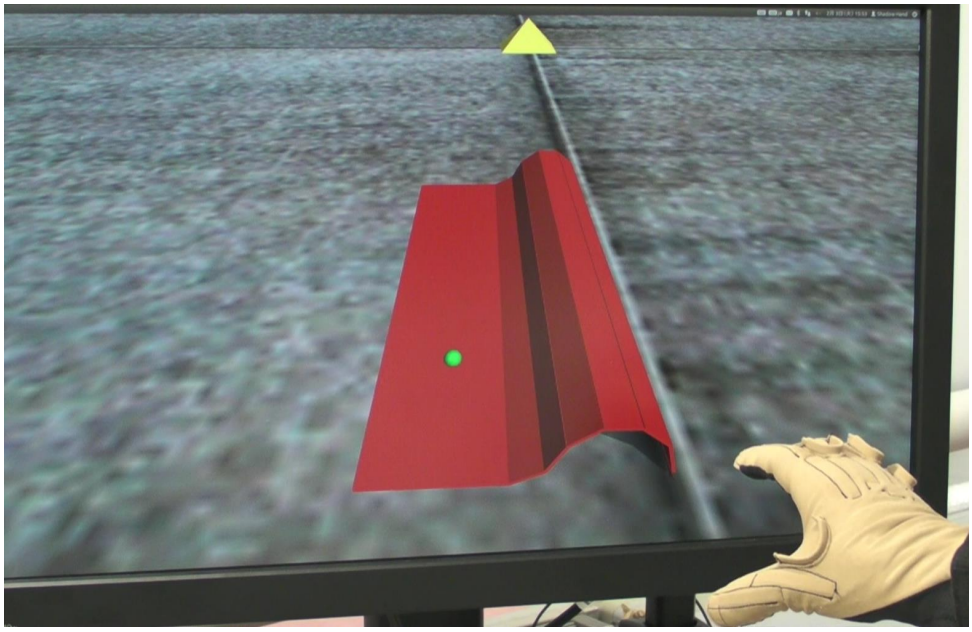


図 17 実験風景

過時間，ページの高さの情報を入力とする放射基底関数 (RBF) を利用した．

$$\boldsymbol{a}^h \quad = \quad \boldsymbol{w}^{\mathrm{T}} \boldsymbol{\phi} \left(\boldsymbol{s}^r, t, \boldsymbol{s}^p \quad \right) \tag{21}$$

5.3 オペレータの行動方策モデルの作成手順

オペレータの行動方策モデルの作成手順を以下に示す．

1. cyberglove を装着した被験者がページめくりタスクを実施する．
2. 物理シミュレーション内の縦方向と横方向の 2 自由度，物理シミュレーション内の指の触覚情報，ページの持ち上がった高さを計測する．
3. 計測の後計測されたデータに対して K 近傍法を用いることで基底関数の平均と分散を算出する．

4. 最小二乗法から，基底関数の重みを決定する．

このタスクは被験者一人に対して 50 回行った．

K 近傍法　K 近傍法では，あるオブジェクトの分類に，k 個の最近傍のオブジェクト群で最も一般的なクラスをそのオブジェクトに割り当てる．本研究では，獲得したデータに対して MATLAB 内の関数 kmeans を利用し，各クラスの平均と分散を算出した．

オペレータの行動方策モデルによる再現性の評価　オペレータの行動法方策モデルによって生成される入力が，どの程度被験者の入力の再現しているか確認する．誤差の評価方法として，疑似的に再現した入力と本来のオペレータの入力の差の平均を利用した．K 近傍法を用いた場合の基底関数の個数は 80 個である．この基底関数は，様々なクラス分類数において最も被験者の入力と誤差の少ないものを選択した．また，各入力次元を最大値と最小値の範囲で 4 つに量子化する基底関数を利用した場合と比較する．量子化法を用いた場合の基底関数の個数は 256 個である．

表 1 に，K 近傍法と量子化法の比較結果を示す．結果から，K 近傍法の方が量子化法よりも誤差が少ないことがわかる．

	親指 第 4 関節	親指 第 5 関節
K-mean	6.21[deg]	0.42[deg]
Grid	13.13[deg]	0.50[deg]

表 1　誤差の平均

図 18 に，親指の各関節の追従結果を示す．このグラフは，横軸がタイムステップ [steps]，縦軸が各関節の角度 [deg] である．図 18 中 (a) は親指の第 4 関節の関節角度の推移を，(b) は親指の第 5 関節の関節角度の推移を示す．青いラインが追従する対象となる関節角度の推移，緑のラインが量子化法によって生成された関節角度の推移，赤いラインが K 近傍法によって生成された関節角度の推移であ

る．図 18 から，K 近傍法の方が，量子化法よりも入力の再現性が高いことがわかる．

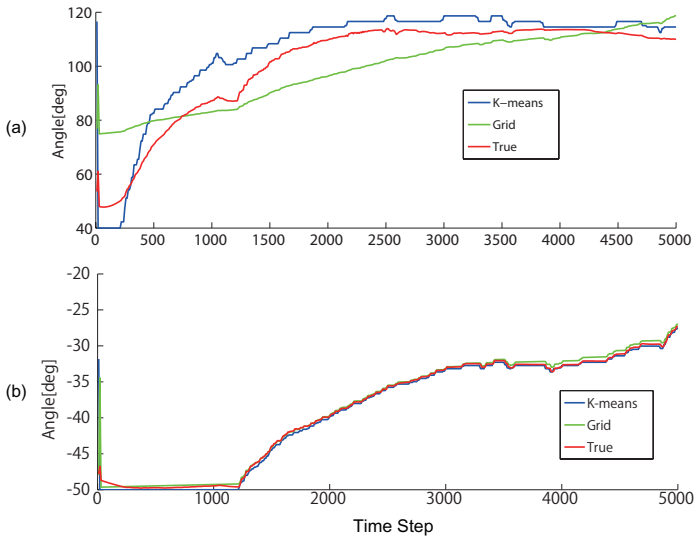


図 18　関節角度の再現性の比較

以上の結果から，K 近傍法を用いた基底関数を利用するオペレータの行動方策モデルが被験者の入力を再現していることと，少ない基底関数で量子化法よりも誤差の少ない入力の生成が行われていることが確認された．

5.4　ページめくり実験

ページめくり実験の詳細を以下に示す．

5.4.1　学習実験

ここでは，器用さ支援方策を学習する実験の詳細について示す．学習の手順は次の通りである．

1. cyberglove を装着した被験者がページめくりタスクを実施する．

2. 5.2 節の手順に従って，オペレータの行動方策モデルを作成する．

3. オペレータの行動方策モデルが，自律的にページめくりタスクを実施する．

4. EM 方策探索に従って方策パラメータ $\boldsymbol{\theta} = (\boldsymbol{w}, \boldsymbol{\sigma})^T$ の更新を行う．

5. 方策パラメータの収束判定を行い，収束していれば終了．収束していなければ手順 3 に戻る．

被験者は，右利きの成人男性 1 名である．各 Iteration での試行回数は 100 回に設定した．学習による期待報酬の推移を確認し，EM 方策探索による学習過程を確認する．また，方策の学習結果を確認するために，各 Iteration において学習された方策を用いて挙動の比較を行う．

5.4.2 被験者によるページめくり実験

ここでは，5.4.1 節で学習された支援方策を用いて行った被験者によるページめくり実験の詳細について示す．今回行った実験を以下に列挙する．

- 作成したモデルと同一の被験者によるページめくり実験
- 作成したモデルと異なるモデルを持つ被験者によるページめくり実験
- 作成したモデルと同一の被験者による複数の動作によるページめくり実験

まず，被験者は右手に cyberglove(5.1.1 節) を装着し物理シミュレーション (5.1.2 節) の操作を練習する．この際物理シミュレーション上にはページのみを表示し，ページめくりタスクの練習を行うことはできないものとする．

次に，物理シミュレーションを用いたページめくりタスクを 1 回 30 秒間で 30 回実施する．この時各試行ごとに物理シミュレーション環境をリセットし，ページや物理シミュレーションシミュレータ上の指の初期位置に変化はないものとする．物理シミュレーションの初期状態を図 19 に示す．

また，被験者によって実施されるページめくり動作は 1 回の試行につきページめくり動作を 1 回に制限する．被験者は健常な成人男性 5 名である．各被験者の詳細を以下に示す．

図 19 初期状態

被験者 A オペレータの動作モデル作成に参加し，かつページめくりタスクに慣れた被験者

被験者 B，C ページめくりタスクに慣れた被験者

被験者 D，E ページめくりタスクに慣れていない被験者

この時の報酬の平均と器用さ支援方策を適用しない時の報酬の平均を比較し，提案手法の有効性を確認する．

次に，被験者 A によって以下の動作を実施する．

1. 素早い動作．
2. ゆっくりとした動作．

3. 1 度ページを持ち上げた後にロボットハンドを初期位置に戻し，再びページをめくる動作.

各動作中のページのめくられた高さの確認を行い，速度や動作の異なるオペレータからの入力に対する提案手法の有効性を確認する.

5.5 実験結果

ページめくり実験の結果を以下に示す.

5.5.1 学習実験

ここでは，支援方策の学習結果についての結果を示す．図 20 に Iteration 毎の期待報酬の推移を示す．実線は 3 回の学習実験で得られた期待報酬の平均を，エラーバーはその標準偏差を表す．図 20 から，報酬値が初期値より上昇しており，期待報酬の高い方策パラメータが学習されてることがわかる．また，約 1200 Iteration で学習が収束していることが確認できる.

図 21 に，学習の過程で獲得された方策を用いたページめくり動作の挙動を示す．図 21 から，1 Iteration 目ではページに触れない試行が，10 Iteration 目ではページに触れるもののページがめくれない試行が，300 Iteration 目ではページがめくれる試行が，1200 Iteration 目では 300 Iteration 目よりページがめくれる試行が確認された．この結果から，ページめくりを支援する方策が学習されていることが確認できる.

5.5.2 被験者によるページめくり実験

ここでは，被験者によるページめくり実験についての結果を示す．図 22 にページめくりタスクの実験結果を示す．青いヒストグラムが器用さ支援がある時，赤いヒストグラムが器用さ支援がない時の結果である.

被験者 A，B，C のタスク達成度は器用さ支援の有無に依らず高く，被験者 D，E のタスク達成度は器用さ支援の有無に依らず低い結果となった.

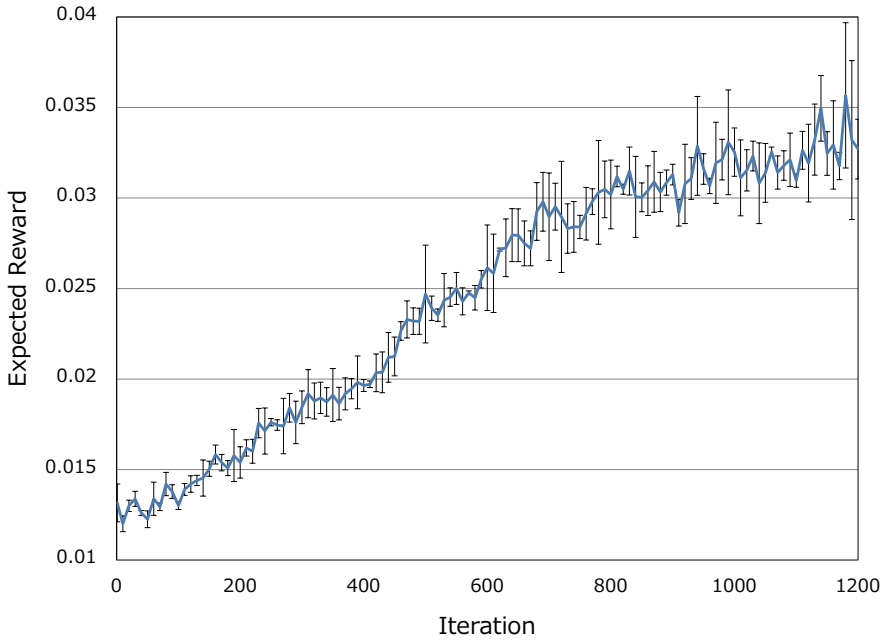


図 20　期待報酬の推移

このような差が発生する要因として次のようなことが考えられる．被験者 A，B，C のタスク達成度が高い理由として，被験者個人の器用さやページめくりタスクへの慣れが要因であると思われる．また，被験者の指の関節の可動域がそれぞれ異なるという点が挙げられる．そのため，cyberglove によって計測される関節角度にオペレータ毎の差が存在する．次に，学習の際に用いられたヒューマンモデルと行動方策が異なるという点が挙げられる．そのため，入力と出力の関係が異なり操作の難易度が上昇する．これらの誤差修正にはキャリブレーションが必要となる．

しかしながら，これらの差が存在する条件下でも器用さ支援がある場合のタスク達成度が上昇していることが確認された．この結果から提案した Shared Control の枠組みでは，オペレータによる不確実性に柔軟に対処できると考えられる．

また，器用さ支援がある時とない時のタスク達成度の差は被験者 A が最も大きい．これは，被験者 A がオペレータの行動モデル作成に参加し，学習がそのモデルに基づいて実施されたためであると思われる．

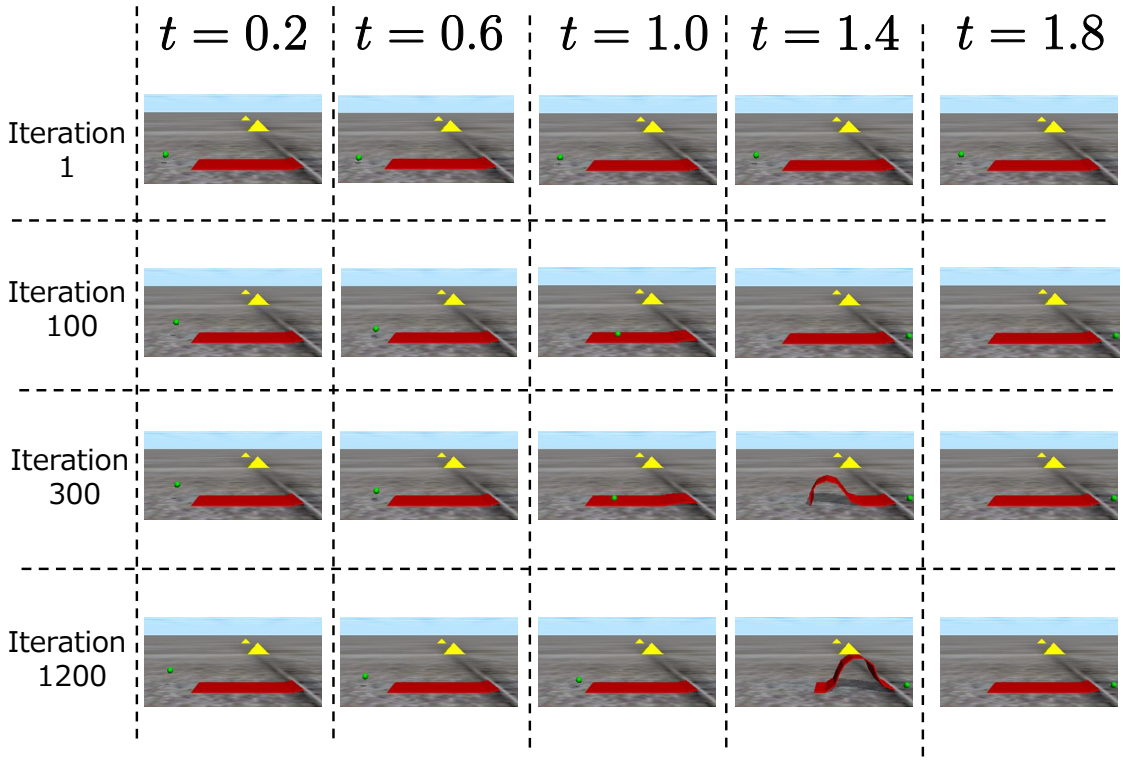


図 21　方策の更新に伴う挙動の変化

図 23 に，被験者 A による複数の動作を行った時のページの高さの時間推移を示す．図 23 から，早い動作，遅い動作，2 度ページをめくる動作等のように，実施する動作に差があってもページがめくられていることがわかる．すなわち，動作の早さやページめくり動作の回数に関係なく，オペレータがページをめくる意思を持つ時に器用さの支援が実施される．このことから，学習された器用さ支援方策はオペレータと協調的にロボットハンドの入力を決定していることがわかる．

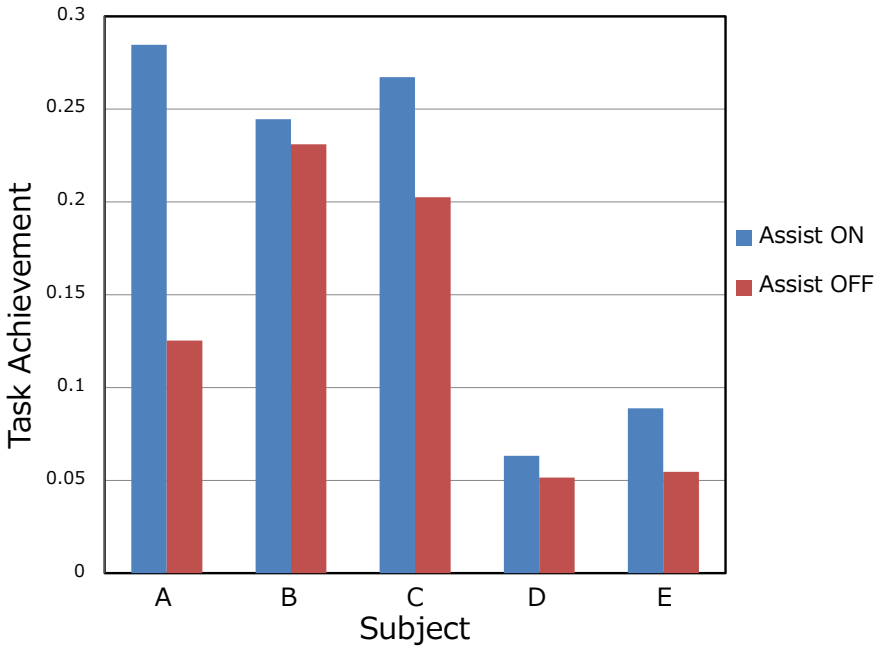


図 22　被験者実験の結果

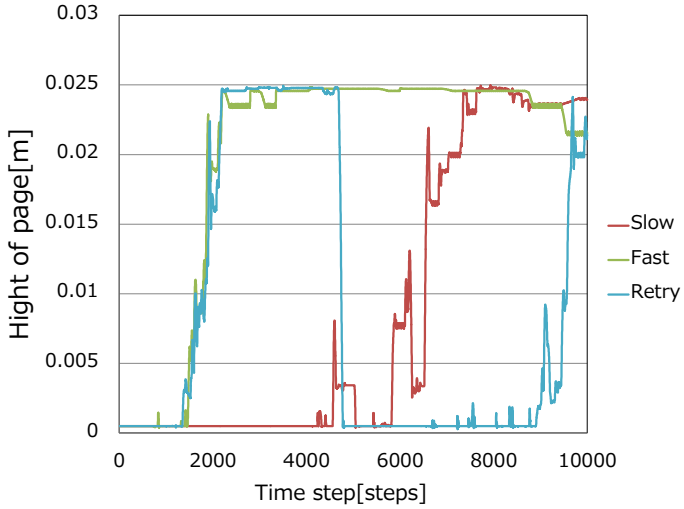


図 23　多様性の検証

6. 結論

ここでは，まとめと器用さ支援の実機実現に向けての課題を示す．

6.1 まとめ

本研究のまとめを以下に列挙する．

1. 遠隔操作型ロボットハンドの器用さ向上のために，触覚情報に基づいて入力を決定するシステムを構成した．
2. オペレータや環境のモデル化が困難であるため，これらのモデル化を行わずに Shared Control を設計するアプローチを提案した．本研究では，POMDP によってシステムをモデル化しオペレータとロボットと環境の相互作用から方策を自律的に設計する手法を提案した．
3. 提案手法を用いてページめくりタスクのための器用さ支援方策の設計を行い，学習が進むにつれてページの到達する高さの増加を確認した．
4. 被験者実験を行い，学習した器用さ支援方策によってページめくりタスクの達成度が上昇することを確認した．また，器用さ支援方策がオペレータと協調的に入力を決定することを確認した．

6.2 実機での器用さ支援の実現に向けて

実機実現に向けての課題として複数の項目が挙げられるので以下に列挙する．

1. 実験回数を増やし，学習に利用するオペレータの行動方策モデルと異なるモデルを持つ被験者への挙動の統計学的な解析を行う必要がある．
2. 支援に関する罰則項を調整することによる，学習された方策の挙動の変化を比較する必要がある．

3. 複数の被験者によるモデルを用いて学習を行い，タスク達成度の上昇度の比較を行う必要がある．また，学習に使う方策モデルと被験者のモデルが異なる場合のタスク達成度の上昇率を確認し，統計的な解析を行う必要がある．

4. 本研究では物理シミュレーションによる高速な学習が可能であったが，実機で行う場合非常に時間がかかる．そのため，実機実現の際にはサンプルリユース [17] 等の学習時間の短縮手法が必要であると考える．

5. 現在の設定では，取り扱う実問題に関する報酬関数の設計が必要であるため，試行錯誤的に報酬関数を設計する必要がある．学習時間の短縮のために，逆強化学習法等を持ちいた一般的な枠組みを提案する必要がある．

6. オペレータの入力が不確実性を持つ場合を想定し，その状況下での支援方策の学習と有効性を確認する必要がある．

謝辞

はじめに本研究を進めるにあたり，知能システム制御研究室の杉本謙二教授には研究のご指導に留まらず，研究者，技術者としてあるべき姿勢，物事に対する考え方など，様々な面でご指導，ご鞭撻を頂きました．心よりお礼申し上げます．

また，お忙しい中，副指導教員を引き受けてくださり，ゼミナール等で大変貴重な意見，ご提案を頂きましたロボティクス研究室の小笠原司教授に深く感謝いたします．

松原崇充助教には本研究を進めるうえでの議論に加えて，研究の進め方や論文の書き方に至るまで，貴重なご助言，細部に渡る熱心なご指導を賜りました．改めて厚く御礼申し上げます．

岡山大学 平田健太郎教授には，定例研究会や学生室での雑談において鋭いご指摘や的確なアドバイスを頂き，至らない私をサポートして頂きました．ここに深く感謝致します．

電気通信大学 小木曽公尚准教授には研究に関するご指摘や普段の生活面の方でもアドバイスを頂きました．ここに深く感謝いたします．

南裕樹助教には研究に関するご指摘や研究室での研究環境の整備をして頂きました．ここに深く感謝いたします．

日本バイナリー株式会社の池田潔氏，磯部正利氏には，実験環境を整える上で手厚いサポートをして頂きました．ここに感謝の意を表します．

そして実機の環境構築にあたり，株式会社高取電機工作所の辰巳雅紀氏には機械部品の作成をして頂きました．心より感謝致します．

日々の研究生活において多くの事務手続きを引き受けて頂きました秘書の林英子さんに心より感謝いたします．

研究に関して的確な助言と援助をしていただきました後期博士課程の田中大介さんには心より感謝いたします．

また，研究に留まらず，学生生活においても多大なサポートをして頂いた知能システム制御研究室の諸氏に深く，心より感謝致します．

最後になりましたが，研究，学生生活の両面で様々なアドバイスを頂いた先輩方，研究生活を陰で支えて頂いた両親, ならびに友人に心より感謝いたします．

参考文献

[1] Craig Sherstan and Patrick M. Pilarski. Multilayer general value function for robotic prediction and control. *IROS Workshop on AI and Robotics*, pp. 1 – 6, 2014.

[2] A. Dragan and S. Srinivasa. Assistive teleoperation: A new domain for interactive learning. *AAAI fall symposium on robots interactively learning from human teachers.*, pp. 1 – 4, 2012.

[3] Marayong P., Bettini A., and Okamura A. Effect of virtual fixture compliance on human-machine cooperative manipulation. *IEEE/RSJ International Conference*, Vol. 2, pp. 1089 – 1095, 2002.

[4] Jorn Vogel, Claudio Castellini, and Patrick van der Smagt. Emg-based teleoperation and manipulation with the dlr lwr-3. In *IEEE/RSJ International Conference on Intelligent Robots and systems*, pp. 672–678, 2011.

[5] Hiroshi Yokoi, Alejandro Hernandez Arieta, Ryu Katoh, Wenwei Yu, Ichiro Watanabe, and Masaharu Maruishi. Mutual adaptation in a prosthetics application. *Embodied Artificial Intelligence Lecture Notes in Computer Science*, Vol. 3139, pp. 146–159, 2004.

[6] Susumu Tachi, Kouta Minamizawa, Masahiro Furukawa, Katsunari Sato. テレイグジスタンスの研究 (第 65 報)-telesar5:触覚を伝えるテレイグジスタンスロボットシステム-. エンターテイメントコンピューティング, Vol. 1, pp. 02A–01, 2011.

[7] Eric Rombkas, Cara E.Stepp, Chelsey Chang, Mark Malhotra, and Yoky Matsuoka. Vibrotactile sensory substitution for electromyographic control of object manipulation. *IEEE Transaction on Biomedical Engineering*, Vol. 60, pp. 2226–2232, 2013.

[8] Eric Sauser, Brenna D.Argall, Giorgio Metta, and Aude G.Billard. Iterative learning of grasp adaptation through human corrections. *Robotics and Autonomous Systems*, Vol. 60, pp. 55–71, 2012.

[9] Jun Ueda, Ryoji Negi, and Tsuneo Yoshikawa. Acquisition of a page turning skill for a multifingered hand using reinforcement learning. *Advanced Robotics*, Vol. 18, No. 1, pp. 101–114, 2004.

[10] N. Daoud, J.P. Gazeau, S.Zeghloul, and M.Arsicault. A fast grasp synthesis method for online manipulation. *Robotics and Autonomous Systems*, Vol. 59, pp. 421–427, 2011.

[11] N. Daoud, J.P. Gazeau, S.Zeghloul, and M.Arsicault. A real-time strategy for dexterous manipulation:fingertips motion planning, force sensing and grasp stability. *Robotics and Autonomous Systems*, Vol. 60, pp. 377–386, 2012.

[12] Jean-Philippe Saut and Daniel Sidobre. Efficient model for grasp planning with a multi-fingered hand. *Robotics and Autonomous Systems*, Vol. 60, pp. 347–357, 2012.

[13] O.B. Kroemer, R. Detry, J.Piater, and J.Peters. Combining active learning and reactive control for robot grasping. *Robotics and Autonomous Systems*, Vol. 58, pp. 1105–1116, 2010.

[14] Weston B. Griffin, William R. Provancher, and Mark R. Cutkosky. Feedback strategies for shared control in dexterous telemanipulation. *IEEE/RSJ International Conference on Intelligent Robots and systems*, Vol. 1, No. 1, pp. 2791 – 2796, 2003.

[15] Christian Cipriani, Franco Zaccone, Silvestro Mieera, and M. Chara Carroza. On the shared control of an emg-control prosthetic hand: Analysis of user -prosthesis interaction. *IEEE Transaction on Robotics*, Vol. 24, No. 1, pp. 170–184, 2008.

[16] Jens Kober and Jan Peters. Policy search for motor primitives in robotics. *machine learning*, Vol. 84, No. 1-2, pp. 171–203, 2011.

[17] Hirotaka Haciya, Jan Peters, and Masashi Sugiyama. Reward weighted regression with sample reuse for direct policy search in reinforcement learning. *Neural Computation*, Vol. 23, No. 11, pp. 2798–2832, 2011.