3E3-6

参照点に依存した確率モデルを用いた連続動作の認識と生成

Recognition and Generation of Motion Sequences based on Referece-Point-Dependent Probabilistic Models

杉浦 孔明*1*2　岩橋 直人*1*2
Komei Sugiura　Naoto Iwahashi

*1(独) 情報通信研究機構
National Institute of Information and Communications Technology

*2(株) 国際電気通信基礎技術研究所
ATR

This paper presents a method that recognizes and generates sequential manipulation motions such as "place-on" and "move-away". The method first learns motions by using reference-point-dependent probabilistic models, and then combines transformed probabilistic models. We have conducted physical experiments in which a user demonstrates the manipulations of puppets and toys, and obtained accuracy of 53% for the recognition of sequential motions. Also, the results of motion generation experiments carried out with a robot arm are shown.

1. はじめに

コンピュータビジョンやロボティクスなどの分野において，人間の行動の認識が注目されてきている [Krüger 07]．特にロボットによる見まね学習では，ロボットに新規動作を教示するために，画像からユーザの行動を理解する手法に関する研究が行なわれている．見まね学習は，専門家でないユーザであってもロボットに新規動作を教示できる枠組みとして重要である．

日常的な環境において活動するロボットにとって，「食器棚からコップを取り出す」といった物体を操作する動作は基本的な動作のひとつであるが，このような物体操作概念の学習・認識・生成は簡単ではない．これは同じ「取り出す」動作であっても，物体の配置の違いによって軌道が全く異なるためである．この問題に取り組んだ先行研究として [小川原 01] が挙げられる．[小川原 01] では，操作軌道を 2 物体間の相対軌道として表現し，確率モデルを用いて学習させている．これに対し我々は，動作の基準となる参照点の推定を行ない，参照点に依存した動作の概念を，軌道に関する運動情報 (座標，速度，加速度) の確率モデルとして学習する手法を提案している [羽岡 00, Sugiura 07]．これにより，「回す」のように操作軌道の基準となるような物体が存在しない概念まで含めた物体操作の学習が初めて可能になる．本論文では，このような参照点に依存した動作を結合させることにより，連続した動作を認識・生成する手法について述べる．

2. 参照点に依存した動作の学習

2.1 参照点に依存した動作

空間的移動の概念には，参照点に依存しているものがある．例として，「ユーザが図 1 左図に示す点線に沿って縫いぐるみを動かす」動作を考える．この動作は，参照点を「青い箱」とした場合には，「のせる」というラベルを与えることができる．すなわち，「ぬいぐるみを青い箱にのせた」という説明が妥当である．一方，参照点が「緑色の箱」である場合には，「とびこえさせる」というラベルが適当である．

認知言語学では，外部世界を解釈する主体のプロセスにおいて焦点化される存在のうち，相対的に際立って認知される対象を**トラジェクタ**，これを背景的に位置付けるオブジェクトを**ランドマーク**として区別する [Langacker 87]．これにより，「～の左」や「～から離れた」など，対象の関係に基づく概念を記述している．

連絡先: 杉浦孔明，京都府相楽郡精華町光台 2-2-2, komei.sugiura@atr.jp

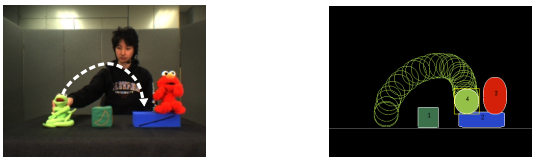


図 1: 左:カメラ画像の例．右:動画像から抽出された観測情報．

ここで，このような参照点に依存する動作の概念を，見まねによりロボットに学習させる問題を考える．このとき，適切な軌道をロボットに再現させるためには，(1) 参照点，(2) 動作固有の座標系タイプ (以下，**固有座標系**), (3) 固有座標系における制御パラメータ，の 3 つを推定しなければならない．例として，「上げる」と「近づける」の概念における，参照点と座標系について考えよう (図 2)．図のように，「上げる」の概念は，トラジェクタの初期位置を参照点とするカメラ座標系を平行移動した座標系を用いることができる．また，「近づける」の概念では，近づく対象をランドマークとし，ランドマークの位置を原点として，ランドマークからトラジェクタへ向かう方向を x 軸とした直交座標系を選ぶのが妥当であろう．

以下では，本論文で提案する確率モデルの結合手法に関する説明の準備として，参照点に依存した動作の学習手法について概説する．

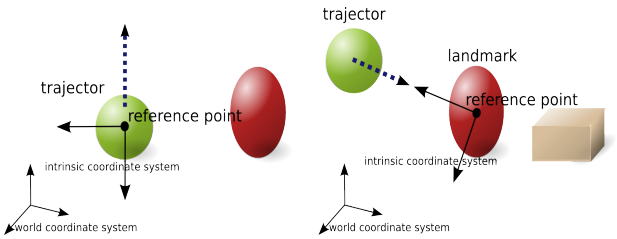


図 2: 動作と参照点・固有座標系の関係．左:「上げる」右:「近づける」

2.2 確率モデルを用いた動作の学習

移動するオブジェクトが一つで複数の静止オブジェクトが存在する動画像が L 個与えられたとする．l 番目の動画像から，移動する物体 (トラジェクタ) の重心軌道 $\mathcal{Y}_l$ と，参照点の候補の集合 $\boldsymbol{R}_l$ を抽出し，観測情報 $\mathcal{V}_l$ とする．つまり，

$$\mathcal{V}_l = (\mathcal{Y}_l, \boldsymbol{R}_l), \tag{1}$$

$$\mathcal{Y}_l = \{\boldsymbol{y}_l(t)|t = 0, 1, ..., T_l\}, \tag{2}$$

$$\boldsymbol{R}_l = \{\boldsymbol{O}_l, \boldsymbol{x}_l(0), \boldsymbol{x}_{\text{center}}\} \triangleq \left\{\boldsymbol{x}^{r_l}\middle|r_l = 1, 2, ..., |\boldsymbol{R}_l|\right\} \tag{3}$$

ここに，$\boldsymbol{y}_l(t)$ はトラジェクタの位置，速度，加速度のベクトル，すなわち $\boldsymbol{y}_l(t) = \left[\boldsymbol{x}_l(t)^\top, \dot{\boldsymbol{x}}_l(t)^\top, \ddot{\boldsymbol{x}}_l(t)^\top\right]^\top$ であり，T_l，$\boldsymbol{O}_l$ はそれぞれ軌道の継続長および静止物体の重心位置の集合である．また，$|\cdot|$ は集合の要素の数を表すとする．

ここで固有座標系 k と，参照点 $\boldsymbol{x}^{r_l}$ によって決定される座標系 $C_k(\boldsymbol{x}^{r_l})$ 上での軌道を ${}^{C_k(\boldsymbol{x}^{r_l})}\mathcal{Y}_l$ と表記することにする．固有座標系は動作概念と 1 対 1 対応であるが，参照点は各学習データ $\mathcal{V}_l$ に対して選択される．ただし，固有座標系は K 種類あり，これらは設計者により与えられる．この設定のもと，参照点インデックス列 $\boldsymbol{r} = \{r_l|l = 1, 2, ..., L\}$・固有座標系タイプ k・軌道に関する確率モデルのパラメータ λ を尤度最大化基準により探索する．

$$(\hat{\boldsymbol{r}}, \hat{k}, \hat{\lambda}) = \underset{\boldsymbol{r},k,\lambda}{\text{argmax}} \sum_{l=1}^{L} \log P(\mathcal{Y}_l|r_l, k, \lambda), \tag{4}$$

$$= \underset{\boldsymbol{r},k,\lambda}{\text{argmax}} \sum_{l=1}^{L} \log P({}^{C_k(\boldsymbol{x}^{r_l})}\mathcal{Y}_l; \lambda), \tag{5}$$

確率モデルとして隠れマルコフモデル (HMM) を用いた場合の上式の解の詳細については，[Sugiura 07] を参照されたい．

3. 参照点に依存した確率モデルの結合

3.1 確率モデルの座標変換

参照点に依存した確率モデルを結合して，連続動作を認識・生成することを考える．HMM による音声認識のように，座標系を共有した複数の確率モデルを結合する手法は広く行なわれている．一方，HMM に基づく音声合成において，学習時に用いた座標系 C と生成時に用いる座標系 C' が異なる場合には，C 上で軌道を生成させた後，C' 上の軌道に変換する．

しかし，参照点に依存した確率モデルの結合の場合には，j 番目の確率モデルのパラメータが $j-1$ 番目の確率モデルのパラメータに依存するため，単一の変換を施すことによって結合することはできない．確率モデルとして HMM を用いるとすると，この変換は図 3 のように模式化できる．

ここで，D 個の固有座標系上の HMM を変換・連結し，世界座標系 W 上の結合 HMM を得ることを考える．ただし，HMM は left-to-right 型であるとし，j 番目の HMM $\lambda^{(j)}$ の状態 s における出力確率密度関数を単一ガウス分布でモデル化する．このとき，状態 s での位置平均ベクトル ${}^{C^{(j)}}\boldsymbol{\mu}_x(s)$ を，固有座標系 $C^{(j)}$ から W への同次変換行列を用いて，以下のように変換する．

$$\begin{bmatrix} {}^{W}\boldsymbol{\mu}_x(s) \\ 1 \end{bmatrix} = \left[\begin{array}{c|c} {}^{W}_{C^{(j)}}R & {}^{W}\boldsymbol{\mu}_x^{(j-1)}(S_{j-1}) \\ \hline \boldsymbol{0} & 1 \end{array}\right] \begin{bmatrix} {}^{C^{(j)}}\boldsymbol{\mu}_x(s) - {}^{C^{(j)}}\boldsymbol{\mu}_x(1) \\ 1 \end{bmatrix} \tag{6}$$

$$(j = 1, 2, ..., D, \quad s = 1, 2, ..., S_j)$$

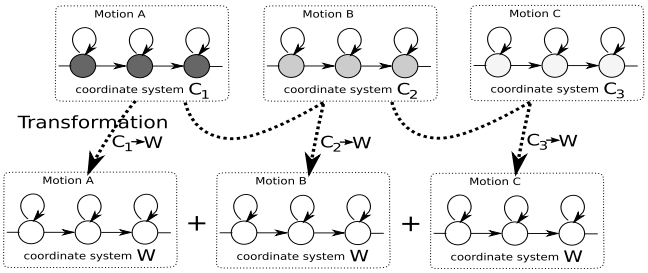


図 3: 参照点に依存した HMM の結合における座標系の変換

ただし，${}^{W}_{C^{(j)}}R$ は $C^{(j)}$ から W への回転行列とし，$s = 0$ は初期状態，$s = S_j + 1$ は最終状態とする．また，速度・加速度に関する平均ベクトル $\boldsymbol{\mu}_{\dot{x}}^{(j)}(s)$，$\boldsymbol{\mu}_{\ddot{x}}^{(j)}(s)$ は ${}^{W}_{C^{(j)}}R$ により回転させる．

一方，位置に関する分散共分散行列 ${}^{C^{(j)}}\boldsymbol{\Sigma}_x(s)$ に関しては，以下の近似を行なう．

$${}^{W}\boldsymbol{\Sigma}_x(s) = {}^{C^{(j)}}\boldsymbol{\Sigma}_x(s) \tag{7}$$

速度・加速度に関する分散も同様な近似により変換を行なう．以上のような近似においては，分散の回転が考えられていない．これは，HMM に基づく軌道生成では分散共分散行列の対角成分のみを用いるのが一般的なためであるが，厳密には非対角成分も考える必要がある．

3.2 結合された HMM による連続動作の認識

連続動作認識は，動詞とそれに対応する固有座標系および HMM パラメータが与えられた下で，軌道 $\mathcal{Y}$ を最も高い確率で出力するモデルを求める問題として定式化することができる．ここで，学習した動詞の集合を $V = \{v_i|i = 1, 2, ..., |V|\}$，動詞 v_i に対応する HMM パラメータを λ_i，動詞 v_i に対応する固有座標系のインデックスを k_i とする．

2.2 節と同様に，移動する物体の重心軌道 $\mathcal{Y}$ と，参照点の候補の集合 $\boldsymbol{R}$ から構成される観測情報 $\mathcal{V}$ が得られたとする．いま，長さ D の動詞-参照点のインデックス列を $(\boldsymbol{i}, \boldsymbol{r}) = (i^{(1)}, i^{(2)}, ..., i^{(D)}, r^{(1)}, r^{(2)}, ..., r^{(D)})$ とする．動詞と HMM パラメータの対応は既知であり，$\boldsymbol{R}$ と参照点のインデックスより参照点の座標が得られるので，前節の方法により結合 HMM $\Lambda_D(\boldsymbol{i}, \boldsymbol{r})$ を得ることができる．このとき，$\mathcal{Y}$ を最も高い確率で出力する動詞-参照点のインデックス列 $(\hat{\boldsymbol{i}}, \hat{\boldsymbol{r}})$ は，以下のようにして求められる．

$$(\hat{\boldsymbol{i}}, \hat{\boldsymbol{r}}) = \underset{\boldsymbol{i},\boldsymbol{r},D}{\text{argmax}}\, P(\mathcal{Y}|\boldsymbol{i}, \boldsymbol{r}, D, \boldsymbol{R}) \tag{8}$$

$$= \underset{\boldsymbol{i},\boldsymbol{r},D}{\text{argmax}}\, P(\mathcal{Y}|\Lambda_D(\boldsymbol{i}, \boldsymbol{r})) \tag{9}$$

3.3 結合された HMM による連続動作の生成

結合された HMM $\Lambda_D(\boldsymbol{i}, \boldsymbol{r})$ を用いて軌道を生成することを考える．いま，静止画像とトラジェクタのインデックス r_{traj} が与えられたとする．2.2 節と同様に，この画像から参照点の候補の集合 $\boldsymbol{R}$ を抽出する．

動詞-参照点のインデックス列 $(\boldsymbol{i}, \boldsymbol{r})$ に対応するトラジェクタの軌道 $\hat{\mathcal{Y}}$ は，以下により得られる．

$$\hat{\mathcal{Y}} = \underset{\mathcal{Y}}{\text{argmax}}\, P(\mathcal{Y}|r_{\text{traj}}, Q_D(\boldsymbol{i}), \boldsymbol{r}, \boldsymbol{R}) \tag{10}$$

$$= \underset{\mathcal{Y}}{\text{argmax}}\, P(\mathcal{Y}|\boldsymbol{x}^{r_{\text{traj}}}, Q_D(\boldsymbol{i}), \Lambda_D(\boldsymbol{i}, \boldsymbol{r})) \tag{11}$$

ただし，$Q_D(\boldsymbol{i})$ は状態系列 (未知) である．[Tokuda 95] で提案されている手法に従って，(11) 式から $\hat{\mathcal{Y}}$ を得る．

次に，目標点 $\boldsymbol{x}_{\text{goal}}$ が与えられたうえで，$\boldsymbol{x}_{\text{goal}}$ に到達する軌道を生成する動詞-参照点のインデックス列 $(\boldsymbol{i}, \boldsymbol{r})$ を求めることを考える．これは，(11) 式右辺の確率を $\boldsymbol{x}_{\text{goal}}$ で条件付けたうえで，$(\boldsymbol{i}, \boldsymbol{r})$ についても探索すればよい．

$$(\hat{\mathcal{Y}}, \hat{\boldsymbol{i}}, \hat{\boldsymbol{r}}) = \underset{\mathcal{Y}, \boldsymbol{i}, \boldsymbol{r}, D}{\operatorname{argmax}} P(\mathcal{Y} | \boldsymbol{x}^{\text{traj}}, \boldsymbol{x}_{\text{goal}}, Q_D(\boldsymbol{i}), \Lambda_D(\boldsymbol{i}, \boldsymbol{r}))$$

4. 実験

4.1 実験設定

実験に用いたロボットシステムは，三菱重工製 PA-10 および Barrett Technology 製 BarrettHand からなる．また，ユーザが物体を操作する軌道は，Point Grey Research 製ステレオカメラ (Bumblebee 2) から得られる画像を処理することで得られる．カメラのフレームレートを 30[frame/sec] とし，解像度を 320x240 とした．図 1 にカメラより得られた画像の例と，それに対応する観測情報の内部表現を示す．

動作認識・生成に用いる動作要素を用意するために，以下の 7 個の概念を学習させた．これらの動作要素はユーザがオブジェクトを操作することにより，教示を行なった．

「上げる」,「近づける」,「離す」,「回す」,「のせる」,
「下げる」,「飛び越えさせる」

固有座標系は以下の C_1 から C_4 のいずれかとする．C_3，C_4 では参照点が唯一に定まるので，参照点の探索を行なわない．また，カメラ座標系と世界座標系の変換は固定であるとする．

C_1: ランドマークを参照点とする，カメラ座標系を平行移動した座標系．ただし，変換後の座標系において $\boldsymbol{x}_l(0)$ の x 座標が負になる場合には，さらに x 軸を反転させる．

C_2: ランドマークを参照点とし，$\boldsymbol{x}_l(0)$ に向かう軸を x 軸とする直交座標系．

C_3: $\boldsymbol{x}_l(0)$ を参照点とするカメラ座標系を平行移動した座標系．

C_4: $\boldsymbol{x}_{\text{center}}$ を参照点とするカメラ座標系を平行移動した座標系．

探索の深さに関しては，D の最大値を 3 とする．また，軌道の生成においてオブジェクトとの衝突は考えない．図 4 に学習データの例と学習の結果選択された固有座標系を示す．

4.2 実験 (1): 連続動作の認識

動詞列と参照点を自然言語でユーザに呈示し，物体を操作させる．図 5 上段は，ユーザにそれぞれ「物体 1 を物体 3 にのせた後，物体 2 にのせる」「物体 1 を回した後，物体 2 から離す」ような指示を呈示した場合に得られた軌道を示している．6 種類の動詞列をランダムに選んだうえで，物体の配置を変化させて 5 種類の環境でユーザに物体を操作させた．すなわち，評価に用いるテストセットのサイズは 30 である．

図 5 下段に，上位 3 位までの認識結果と対数尤度を示す．図において，認識結果の動詞列は＜動詞，物体 ID ＞の列により表されている．図より，図 5 上段左図の軌道に対しては，正しい認識結果が得られていることがわかる．一方，図 5 上段右図に対しては，1 位の認識結果は不正解である．以下で定量的に示すように,「近づける」「離す」など固有座標系が C_2 である動詞を含む認識において，認識率が低くなる傾向が見られた．

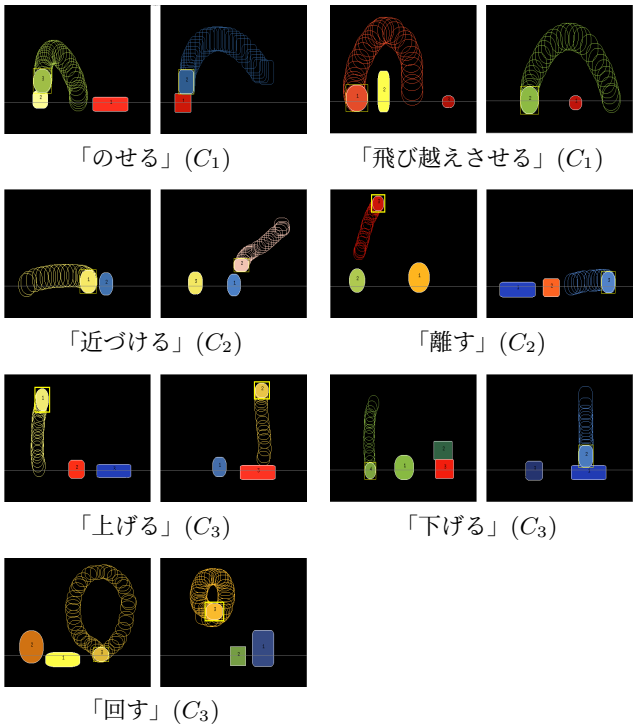
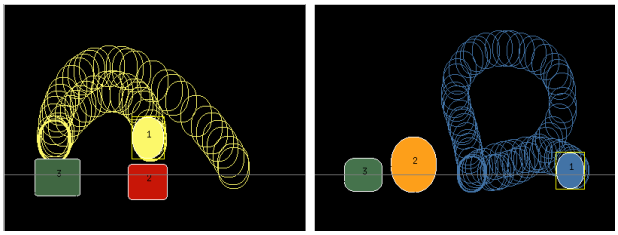


図 4: 学習データの例

これは，(7) 式における分散の近似に起因すると考えられる．用意した固有座標系のうち C_2 のみが分散の回転を必要とするが，提案手法は分散の回転を考慮していない．

表 1 に各テストセットの正解数と認識率を示す．表 1 において，n-best で示される列は上位 n 位までに正解が含まれる数を示す．1-best，2-best，3-best の認識率はそれぞれ，53%，80%，86%であった．ここで，固有座標系タイプが C_2 である動詞を含む列，すなわち (2)，(5)，(6) の認識 (1-best) に注目すると，認識率は 27% (4/15) である．一方，C_2 の動詞を含まない (1)，(3)，(4) の動詞列の認識率は 80% (12/15) である．このことからも，分散を回転する必要がある C_2 の動詞の認識率が低いことがわかる．



1. ＜のせる，3 ＞＜のせる，2 ＞: -22.04
2. ＜飛び越えさせる，2 ＞＜のせる，2 ＞: -23.79
3. ＜のせる，3 ＞＜離す，3 ＞: -28.79

1. ＜離す，2 ＞: -22.18
2. ＜回す，1 ＞ ＜離す，2 ＞: -22.65
3. ＜回す，1 ＞: -25.06

図 5: 物体操作軌道と認識結果の例

4.3 実験 (2): 連続動作の生成

提案手法により生成された軌道の例を図 7 に示す．図 7 は，トラジェクタとして物体 2，動詞列として＜離す,1 ＞＜飛び越えさせる,4 ＞＜近づける,4 ＞を入力とし，得たものである．図において，実線は動詞列指定モードにおいて生成された軌道で

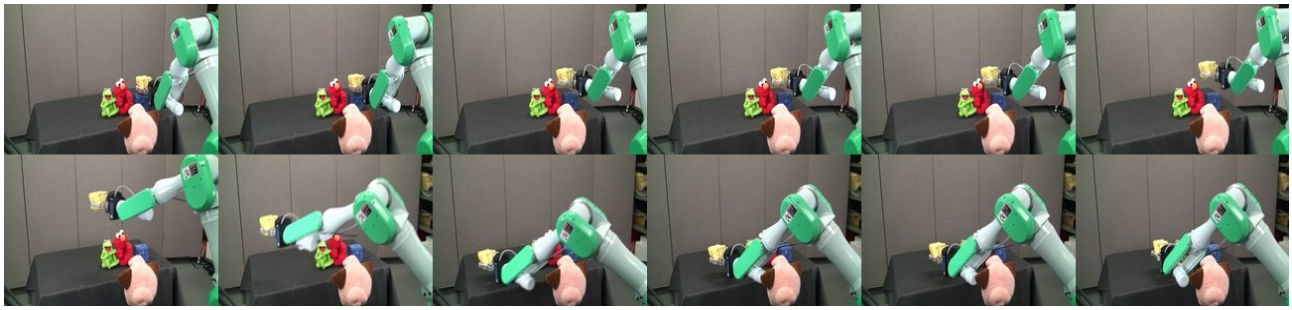


図 6: 図 7 に示す軌道に沿った物体操作

表 1: 各テストセットにおける正解数および認識率

動詞列	1-best	2-best	3-best
(1) 回す+回す	5	5	5
(2) 離す+近づける	0	3	4
(3) のせる+のせる	3	5	5
(4) 回す+飛び越えさせる	4	4	4
(5) 回す+離す	2	4	5
(6) 近づける+のせる	2	3	3
計	16 (53%)	24 (80%)	26 (87%)

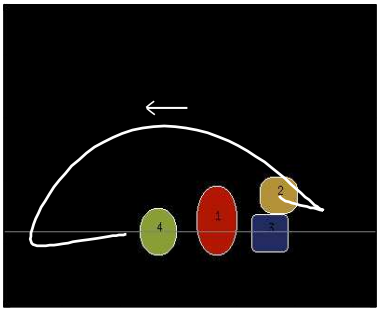


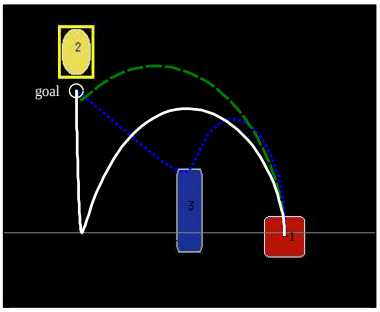
図 7: 動詞列指定モードにおける生成された軌道の例

ある．図 7 の軌道は,「黄色い箱 (ID:2) をエルモ (ID:1) から離した後，カーミット (ID:4) の上を飛び越えさせてから，カーミットに近づける」動作を意味している．図 6 の連続写真は，この軌道をロボットアームに実行させたものである．ここで，図 6 における時間の流れは，左上段→右上段→左下段→右下段の順である．図 7, 図 6 より提案手法が適切な軌道を生成していることがわかる．

目標点指定モードにおける結果の例を図 8 に示す．図 8 は，物体 1 をトラジェクタとし，図中の目標点 (“goal”) までの軌道を生成させた例である．ただし図には，対数尤度の上位 3 位までの動詞列から得た軌道をそれぞれ実線，破線，点線で示した．図より,「物体 3 を飛び越えさせてから，物体 2 に近づける」動作と解釈できる軌道が生成されていることがわかる．

5. おわりに

本論文では,「のせる」「回す」など物体を操作する動作を，参照点に依存した確率モデルを用いて学習した後，確率モデル



1. 実線 ＜飛び越えさせる，3 ＞＜近づける，2 ＞: -16.45
2. 破線 ＜飛び越えさせる，3 ＞: -18.66
3. 点線 ＜のせる，3 ＞＜近づける，2 ＞: -24.00

図 8: 目標点指定モードにおける生成された軌道の例

を結合させて連続動作を認識・生成する方法について述べた．本手法の応用としては，作業者の動作の理解のための画像認識などが挙げられる．

謝辞

本研究は，日本学術振興会科学研究費補助金 (基盤研究 (C) 課題番号 20500186) および立石科学技術振興財団による研究助成を受け実施したものである．

参考文献

[Krüger 07] Krüger, V., Kragic, D., Ude, A., and Geib, C.: The meaning of action: a review on action recognition and mapping, *Advanced Robotics*, Vol. 21, No. 13, pp. 1473–1501 (2007)

[Langacker 87] Langacker, R. W.: *Foundations of Cognitive Grammar: Theoretical Prerequisites*, Stanford Univ Pr (1987)

[Sugiura 07] Sugiura, K. and Iwahashi, N.: Learning object-manipulation verbs for human-robot communication, in *Proceedings of the 2007 workshop on Multimodal interfaces in semantic interaction*, pp. 32–38 (2007)

[Tokuda 95] Tokuda, K., Kobayashi, T., and Imai, S.: Speech parameter generation from HMM using dynamic features, in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*, pp. 660–663 (1995)

[羽岡 00] 羽岡 哲郎, 岩橋 直人：言語獲得のための参照点に依存した空間的移動の概念の学習, 信学技報, PRMU2000-105, pp. 39–46 (2000)

[小川原 01] 小川原 光一：注視点に基づく手作業の理解とそのロボットへの実装に関する研究, PhD thesis, 東京大学 (2001)