カーネル正準相関分析による視覚姿勢情報統合に関する構成論的研究

Constructive Approach to integration of Visual and Posture Information Using Kernel Canonical Correlation Analysis

○谷口　忠大，　　太田　敬太
○Tadahiro Taniguchi，　Keita Ota
立命館大学
Ritsumeikan University

Abstract: In this paper, we propose a novel constructive computational model which integrates visual information and joint angular information about human or humanoid robot ’s posture. It is said that human brain has body schema which represents body image and involves sensory-motor information efficiently. When a subject imitates another person ’s behavior, the imitator has to map the visual image of the observed posture to the imitator ’s joint angular. This is not easy to extract joint angular information from raw visual image. In this paper, we introduce kernel canonical correlation analysis (KCCA) which can extract a common hidden continuous variable from two different information sources. We apply the method to two-link robot arm in a simulation environment, and evaluate the effectiveness in an experiment.

1　はじめに

人間の知能を理解する上で，模倣学習の理解は本質的である．筆者らは連続的な動作提示からロボットが模倣学習を行うための非分節動作模倣の枠組みを提案してきた [1, 2]．また，人間とロボットが単純なインタラクションを通じて役割交代模倣を行う枠組みを，生成的な確率モデルとして表現し EM アルゴリズムを通して役割の推定と同時に応答戦略の学習を行う枠組みを提案した [3]．しかし，これらの模倣学習はすべて，動作を生成する際の関節角情報や動作の種類が観測可能とした定式化になっている．視覚に基づく観察からの連続的相互作用に基づく模倣学習を実現するためには，いかに観察した他者の動作を自らの身体に動作に対応づけるかという問題がもっとも基本的な問題として存在する．

1.1　観察した身体と自らの身体との対応

一方で，人間は他者を模倣する際に他者の動作を自らの身体の上の動作へと写像することができる．これは模倣学習を実現する上で非常に基本的な機能である．人間の脳においてはミラーニューロンがこのような他者の身体動作と自己の身体動作のマッピングを行うとされるが，ミラーニューロンの担う計算処理がいかに脳内にボトムアップに形成されるかは明確ではない [4]．Nehaniv らは模倣学習において観察した他者の関節を自らの身体の関節にいかにマッピングするかという問題を correspondence problem と呼び，多くの組み合わせの可能性があることを指摘した [5]．しかし，実際には，その関節角同士の対応についての組み合わせにという問題に至る前には，他者の動作を視覚像を通して解釈する際にリンク系である事を前提として関節角情報へデコードする必要がある．一方で，動作情報と視覚情報を相互制約的に利用することができるような効果的な情報抽出とその低次元表現か可能であれば，模倣学習の上でも，動作生成の上でも有用である．本稿ではマルチモーダル情報の間に非線形な関係性を見いだし情報抽出を行うことが可能なカーネル正準相関分析 [6] に着目し視覚姿勢感覚統合のための構成論的な

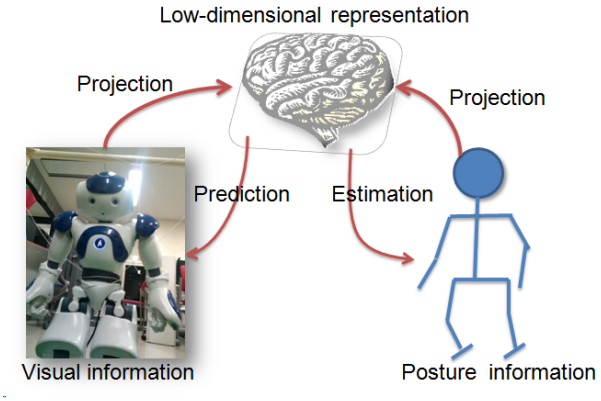

図 1: 視覚情報と姿勢情報の統合

アルゴリズムを検討する．

1.2　身体図式の獲得

また，人間は自らの身体図式という低次の表現を脳内に持つことで，効率的に自らの動作を形成すると考えられるが，このような動作の低次元表現をいかに得るかという問題も重要な問題となる [7]．人間やサルが自らの身体の表現，つまり身体図式を脳内に持っていることは明らかになってきている [7]．入来らは道具を持った際に身体図式が道具の先端まで拡張されることを示した [8]．鍋島らはこの身体図式の拡張を表現する構成論的なモデルを提案している [9]．また，運動野や体性感覚野には各身体部位に強く相関したニューロン群があることが知られており，その部位の上に描かれた人の姿のマップはペンフィールドのホムンクルスと呼ばれている．Lars らは AIBO の各センサー・モータ情報の相互情報量に応じて配列を変える事によって，センサ・モータの情報を二次元マップ上に自己組織的に並び替えペンフィールドのホムンクルスのような，マップを形成することができることを示した [10]．類似の研究は，杉浦らによってもなされている [11]．それぞれ相互情報量や相関をみる，情報の遅延をみるなど手法に違いはあるが，各センサ・モータ情報から各素子を自己組織的に低次元マップ上に再配置することで身体図式を獲得しようという点では共通している．

模倣学習を目的とすると，如何に他者や自己の身体の視覚像という外部の情報から関節角度といった内部の情報を生成するかという問題が身体図式に関わる大きな問題となる．この視点から，本稿では身体図式とは姿勢や視覚像を統合する脳内の身体に関する低次元表現ととらえ，これをロボットにボトムアップに形成させることを目指す．

2　カーネル正準相関分析

本章では視覚情報と姿勢情報を統合するために本稿で用いるカーネル正準相関分析について概説する [6]．通常，複数のチャネル，モダリティから得られた情報の中にはお互いに関係し問題の認識に関係するものも含まれれれば，無関係なものも含まれている．たとえば，画像についは光の加減や不要な背景など認識の対象とは無関係な情報によって大きく変化することがある．本稿では，身体動作を視覚的に観察し，ここから身体動作に関わる成分だけを抜き出すのが課題となる．

このような問題に対しては正準相関分析が有効であると考えられる．正準相関分析 (CCA: Canonical Correlation Analysis) は複数の情報源に共通して含まれる

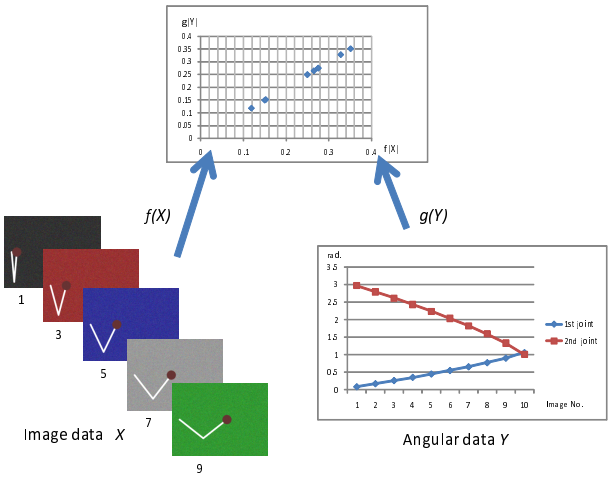

図 2: 正準相関分析の概観

情報だけを抽出することによって，情報を統合するための次元圧縮の手法である．正準相関分析では各モダリティから得られた状態ベクトル x, y に基づいて

$$\rho_n = Cor_n[f(x), g(y)] \tag{1}$$

の最大化を目的とする．つまり，二つの多次元の状態空間 X, Y を考え，これらからスカラー値への写像の組 f, g を考える．n 個の状態ベクトルの組 $(x_i, y_i)(i = 1, 2, ..., n)$ についてこれらの f, g 写像先において相関係数を最大化する f, g を求めるのが正準相関分析である．通常の正準相関分析では関数 f, g に線形性を仮定することで固有値問題に帰着させることができる．しかし，視覚情報や姿勢角情報といった非線形性を有する実データを扱う際には，線形性の保証が困難となる．一方で完全に非線形性を仮定するとこの最適化問題を解くこと自体が困難となる．カーネル正準相関分析ではこれを特徴ベクトルの線形関数に拡張する．つまり

$$f(x) = a^T \Phi_x(x) \tag{2}$$

$$g(y) = b^T \Phi_y(y) \tag{3}$$

と仮定する．しかし，このままで記述力が高すぎるので，正則化項 $\zeta_x \|a\|^2 + \zeta_y \|b\|^2$ 評価関数に付加し過学習の回避を行う．

カーネル関数を用いることで

$$f(x) = \sum_{i=1}^{n} \alpha_i k_x(x^{(i)}, x) \tag{4}$$

$$g(x) = \sum_{i=1}^{n} \beta_i k_y(y^{(i)}, y) \tag{5}$$

と定義される．ここで k_x, k_y は特徴ベクトルの内積で定義されるカーネル関数である．ここで α_i, β_i をそれぞれ並べたものをベクトル α, β と定義する．$x^{(i)}, y^{(i)}$ はそれぞれ i 番目のサンプルデータを示す．また，与えられた x, y のサンプルデータに対する，カーネル関数の値をグラム行列として並べたものをそれぞれ K_x, K_y とする．ここで J_n を単位行列の各成分から $\frac{1}{n}$ を差し引いたものとすると，カーネル正準相関分析の解は一般化固有値問題

$$\begin{pmatrix} 0 & K_x J_n K_y \\ K_y J_n K_x & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \lambda \begin{pmatrix} K_x J_n K_x + \zeta_x K_x & 0 \\ 0 & K_y J_n K_x \quad K_y J_n K_y + \zeta_y K_y \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \tag{6}$$

の解となる．λ は特徴空間上での正則化項付きの相関係数に等しいので，上記の固有値の内できるだけ大きな λ を選ぶことが望ましい．ここで最大の λ を第一正準相関係数と呼ぶ．これらの上位 d 個をとり，それらに対応する d 個の固有ベクトルによって写像を構成することで，X, Y 両状態空間における分布とその関係性を考慮した d 次元空間への低次元化写像が得られる．たとえば，第一正準相関係数に対応する写像により，得られる値を第一正準変量と呼ぶ．本稿では，第一正準変量のみに着目し，視覚・姿勢情報の統合に対するカーネル正準相関分析の有効性について検討する．

3 実験

3.1 実験条件

本研究では目標としては 1 に示したような全身姿勢情報と視覚像との統合を目指すが，その予備的研究として 2 リンク系によるシミュレーション空間上での実験を行う．まず，10 枚の 2 リンク系を上から見た画像を作成した．画像は 64×48 とし，各リンクの長さは 20 とした．左端中央に付け根があり，手先を横軸方向に 5 の位置から 40 の位置まで 10 フレームに渡って移動させた．また，現実系における背景の変化を擬似的に表現するために背景に変化を持たせた．各ピクセルにはランダムに RGB 値に各 [0, 100] の一様乱数値が代入される．これにそれぞれ (0, 0, 0), (100, 100, 100), (0, 0, 100), (100, 0, 0), (0, 100, 0) を RGB 値に足し込むことで Dark, Gray, Red, Blue, Green の 5 種類の背景を準備した．これらを順に切り替えながらフレームを進めることで図 3 に示すような 10 枚の画像を得た．また，テストデータとして，各背景の画像を同様に 10 フレーム別途作成しておく．これらの画像はグレースケール化した上で

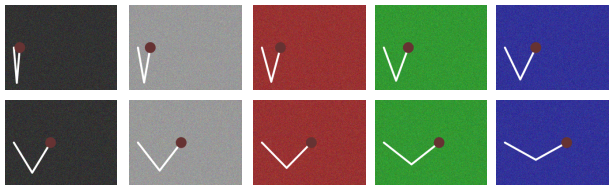

図 3: 学習に用いた画像データ

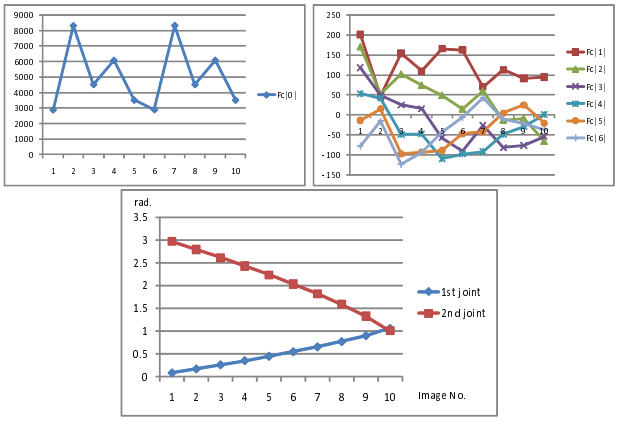

図 4: 各フレームに対する状態ベクトルの変化　上：画像の DCT 係数 下：関節角

離散コサイン変換し，視覚入力ベクトルとして，また，二つの各関節角情報を姿勢入力ベクトルとして学習器に与えた．このときの状態ベクトルの変化を図 4 に示す．画像の DCT 係数については $64 \times 48 = 3072$ 次元データであるが，低周波成分から 7 つの値のみ示している．画像については背景の変化に対応している直流成分が他の次元の値に比べて圧倒的に大きな値を示している事がわかる．故に主成分分析のような手法では直流成分が強く効きすぎてしまい，この画像データから手先位置の変化に着目した低次元情報の抽出を行うことはできない．また，交流成分についてもノイズ成分を含んでいることもあり，非線形性と揺らぎの大きなデータとなっていることがわかる．そこで，カーネル正準相関分析を用いて姿勢の変化と相関の高い情報抽出を行う必要性がでてくる．

カーネル正準相関分析のパラメータは正則化項 $\zeta_x = \zeta_y = 1.0 \times 10^{-2}$ とした．また，カーネルにはガウスカーネル

$$k(x, y) = \exp(-\beta \|x - y\|^2) \tag{7}$$

を用いて，β については画像側で 1.0×10^{-7}, 関節側で 1.0×10^{-2} とした．これらの値は本稿の実験では発見

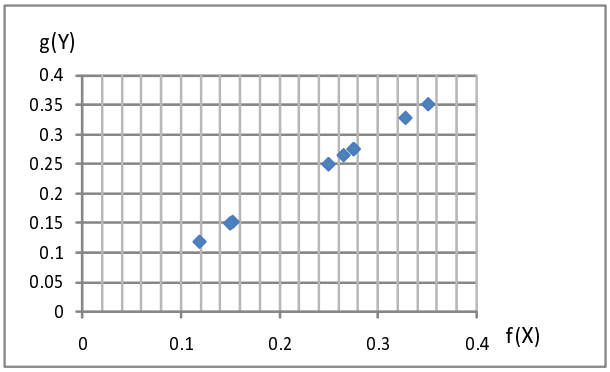

図 5: 写像先の空間での分布

的に決定した[1].

3.2 実験結果

本実験では，サンプルデータを生成した際に潜在連続変数として設定しているのが手先位置という一次元変数であることより，第一正準相関係数および，第一正準変量のみに注目する．得られた第一正準変量を $f(x), g(y)$ について散布図にプロットしたものを図 5 に示す．非常に高い相関が得られていることがわかる．この時の 10 フレームのサンプル画像データに対する第一正準変量の変化を示したのが図 6 の上図である．多少の揺らぎはあるが，手先が前方に進むに従って徐々に増えている事がわかる．手先位置との相関係数は 0.92 であった．しかし，このデータはあくまで学習サンプルデータに対するものであり，過学習を起こしているだけで，適切な特徴抽出を行えていない可能性がある．そこでテストデータとして，5 種類の各背景に対して 10 フレームづつの画像を用意してこれらに対して上記学習サンプルデータから求めた $f(x)$ を用いて写像した値を図 6 の下に示す．背景にかかわらず手先位置が進むに従い値が増加している事がわかる．手先位置との相関係数は 0.80 であった．図 6 より Gray 条件で殆ど値が変化しないという現象が起きていた．Gray を除いた 4 条件では相関係数は 0.93 というサンプルデータを超える結果が得られた．これらの結果より，カーネル正準相関分析を用いることで学習器は潜在連続変量としての手先位置を関節角情報と画像情報との情報統合を通じて抽出することができたと言えるだろう．

[1]本稿では発見的に行っているが，自動的に決定する場合にはクロスバリデーション法を用いることができる．

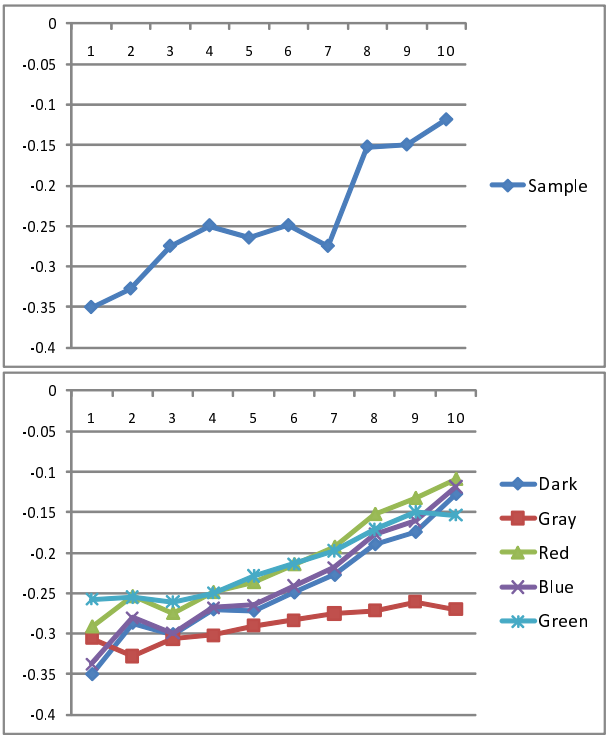

図 6: 第一正準変量の手先位置に対する変化

4 まとめ

本稿では身体の姿勢を映した画像情報について，写し込まれた身体についてのモデル情報を一切持たずに，姿勢角情報との情報統合により，これらの変化に共通した低次元情報を抽出する手法を提案した．カーネル正準相関分析を用いることで潜在連続変量としての手先座標ときわめて高い相関を持つ特徴量を抽出することができた．本研究では背景の変化も考慮しながら，計算機上で仮想的に 2 リンク系を作成し手先位置を変化させながら視覚情報と姿勢情報とを得たが，実画像や実ロボットに対して当該手法が有効であるかどうかは別途検証の必要がある．また，本研究では姿勢情報の側が二次元と，きわめて低次元の情報であった．これに対して人間やヒューマノイドロボットの身体はより高い自由度を持っている．このような条件においても適切な低次元空間を構成できるかどうかも重要な課題である．

5 謝辞

本研究は科学研究費補助金 若手研究（スタートアップ）20800060「非分節な人間機械相互作用を通じた自己組織化型模倣学習機構の構築」，科学研究費補助金学術創成「記号過程を内包した動的適応システムの設

計論」19GS0208 及び国立情報学研究所共同研究助成「言語の超越性と記号創発に関する構成論的研究」の一部支援を受けた.

参考文献

[1] 谷口忠大, 岩橋直人. 複数予測モデル遷移の n-gram 統計に基づく非分節運動系列からの模倣学習手法. 知能と情報（日本知能情報ファジィ学会論文誌）, Vol. 21(6), pp. 1143–1154, 2009.

[2] T. Taniguchi, N. Iwahashi, K. Sugiura, and T. Sawaragi. Constructive approach to role-reversal imitation through unsegmented interactions. *Journal ref: Journal of Robotics and Mechatronics*, Vol. 20, No. 4, pp. 567–577, 2008.

[3] 谷口忠大, 岩橋直人, 中西弘門, 西川郁子. ヒューマン・ロボットインタラクションを通した役割反転模倣に基づく実時間応答戦略獲得. 第 23 回人工知能学会全国大会 in CD-ROM, 2009.

[4] マルコイアコボーニ. ミラーニューロンの発見 - 「物まね細胞」が明かす驚きの脳科学 (ハヤカワ新書 juice). 早川書房, 2009.

[5] C.L. Nehaniv and K. Dautenhahn. The correspondence problem. *Imitation in Animals and Artifacts*, pp. 41–61, 2002.

[6] 赤穂昭太郎. カーネル多変量解析 - 非線形データ解析の新しい展開 (シリーズ確率と情報の科学). 岩波書店, 2008.

[7] サンドラブレイクスリー, マシューブレイクスリー. 脳の中の身体地図 - ボディ・マップのおかげで、たいていのことがうまくいくわけ. インターシフト, 2009.

[8] A. Iriki, M. Tanaka, and Y. Iwamura. Coding of modified body schema during tool use by macaque postcentral neurones. *Neuroreport*, Vol. 7, No. 14, pp. 2325–2330, 1996.

[9] C. Nabeshima, Y. Kuniyoshi, and M. Lungarella. Adaptive body schema for robotic tool-use. *Advanced Robotics*, Vol. 20, No. 10, pp. 1105–1126, 2006.

[10] L.A. Olsson, C.L. Nehaniv, and D. Polani. From unknown sensors and actuators to actions grounded in sensorimotor perceptions. *Connection Science*, Vol. 18, No. 2, pp. 121–144, 2006.

[11] 杉浦孔明. 自律ロボットにおける形態に基づく解釈系の構築. PhD thesis, 京都大学大学院情報学研究科, 2007.

連絡先

谷口忠大
E-mail: taniguchi@ci.ritsumei.ac.jp