Transient-SOM を用いた手画像命令学習システム

羽 野 ともえ*・呉 本 堯**

小 林 邦 和**・大 林 正 直**

A Hand Image Instruction Learning System Using Transient-SOM

Tomoe HANO*, Takashi KUREMOTO**,
Kunikazu KOBAYASHI** and Masanao OBAYASHI**

This paper presents a practical hand image instruction learning system for partner robots using an advanced self-organizing map “Transient-SOM”. The construction of the proposal system is designed with 4 layers: a feature map to classify input images, an action map to acquire correct actions using a reinforcement learning algorithm, a memory layer to preserve “Best Matching Unit (BMU)” after one instruction training is finished, and a feeling map to express robot’s internal state during learning. Experiments using an entertainment robot, AIBO (ERS-7, Sony Co., 2003), showed high effectiveness of the proposed system.

Key Words: hand recognition, instruction learning, human-robot communication, SOM

1. はじめに

人間とロボットのインタラクションやコミュニケーション研究分野において、これまでに、手の形状 (hand shape, gesture, または, posture) や手話 (sign language) によって、人間の意図をマシンに入力しようとする研究は多く行なわれていたが、実用化レベルまでには至っていない^{1), 2)}。その原因は肌色領域の抽出³⁾、3次元空間中の手形状の複雑さ²⁾、命令教示者側 (人間) と行動学習側 (ロボット) の意図理解や共感を実現する困難さなど¹⁾が挙げられる。

本研究ではロボットが利用者から提示された手形状を判別し、人間の意図に従って行動の学習や再学習ができる実用性を重視する手画像命令学習システムを提案する。提案システムでは、HSV 色空間と RGB 色空間の閾値処理を同時に利用する肌色抽出方法を用いて画像中の手領域を検出し、自己組織化マップ (SOM)^{5), 6)}をベースとした特徴分類と命令学習といった複数の機能層をもつ Transient-SOM を構築する⁴⁾。Transient-SOM の学習方式は機能層によって競合学習および学習ベクトル量子化 (LVQ) や強化学習アルゴリズムによって行なわれる。また、ロボットの命令習得状況を感情マップで反映し、感情値をロボットの動作や顔の表情で表現する機能を備える。犬型のエンターテインメントロボット AIBO (ERS-7, ソニー (株))⁸⁾を用いた実機実験で、提案システムの有効性を確認する。

2. 提案システムの構成

提案する手画像命令学習システムのアーキテクチャーを Fig. 1 に示す。

2.1 画像処理

カラー画像から人間の顔や手などの素肌部分を抽出する。これまで、HSV, YCC, RGB, YIQ, XYZ などの表色系に着目した多くの研究が挙げられるが、背景の環境や光源の違いが複雑のため、肌色領域を抽出する方法は確立されていない³⁾。本研究では、HSV 系の H と S の値および RGB 系の R の値を同時に閾値処理することによってカラー画像中の肌色部分を抽出する手法を考案・採用する。われわれは文献³⁾による肌色の HS 分布および室内照明条件を検討し、蛍光灯照明の室内で CCD カメラを用いて取得した 100 枚の実画像を用いて、S, H, R の具体的な閾値を定めた (紙面上の都合、詳細は略)。上記の処理を行なうため、カメラで取得した YCrCb 画像を RGB 画像に、RGB 画像を HSV 画像に変換する必要があり、抽出した部分に対して、ラベリング・平行移動・回転・拡大の処理も行なう。

2.2 Transient-SOM

従来の SOM^{5), 6)}を入力データを分類する一時的な処理層とし、記憶層・行動マップを導入して用い、学習データの種類が増加しても対応できる学習システム Transient-SOM を構築する⁴⁾。

Transient-SOM (Fig. 1 の点線部分) は、特徴ベクトルによって構成する入力層 (Input layer)、2次元 SOM の出力による特徴マップ層 (Feature map)、特徴の異なる入力に対応する行動マップ層 (Action map)、および学習終了時の入力分類・対応行動の情報を保存する記憶層 (Memory layer) という多層構造をもつ。Transient-SOM による画像分類および行動学習のアルゴリズムを以下に示す。

* 山口大学大学院理工学部研究科 (現富士通株式会社)

** 山口大学大学院理工学研究科 宇部市常盤台 2-16-1

* Graduate School of Science and Engineering, Yamaguchi University (now, Fujitsu Ltd.)

** Graduate School of Science and Engineering, Yamaguchi University, 2-16-1 Tokiwadai, Ube

(Received July 26, 2007)

Fig. 1 A learning system with Transient-SOM

• **step 1:** 各マップの初期化. 2次元の特徴マップ (Feature map) に対して, n 次元ベクトルからなるノード $m_i(t)$ にランダムな値を与える ($i = 1, 2, \dots, N \times M, t = 0$). 特徴マップの各ノードの出力を受ける行動マップ (Action map) の各ノードには, 行動番号 a_k をノードのラベルとしてランダムに張り付ける ($k = 1, 2, \dots, K, K < N \times M$).

• **step 2:** 特徴マップの勝者ノード. 勝者ノード (BMU) とは, 入力ベクトルと特徴マップ上のノードのユークリッド距離が最短の場合のノードを指す. すなわち, 勝者ノード C は $C = \arg \min_i \{ \| \mathbf{x} - m_i \| \}$ によって決める. ただし, $\mathbf{x}$ は入力ベクトル ($x_1, x_2, \dots, x_n$), $\| \cdot \|$ はノルム, $i = 1, 2, \dots, N \times M$ である.

• **step 3:** 競合と近傍学習. $m_i(t+1) = m_i(t) + h_{ci}(t)[x(t) - m_i(t)]$ を用いて, m_i を更新する. ただし, 学習回数 $t = 0, 1, 2, \dots, L$, $h_{ci}(t)$ は特徴マップ上の勝者ノードに関する近傍関数である. 計算コスト削減のため, 従来の Gaussian 関数⁵⁾でなく, ここでは $h \propto \| \mathbf{x}(t) - m_i(t) \|$ とする. $0 < h < 1$.

• **step 4:** ラベル付け. 入力ベクトルはまず記憶部に保存された勝者ノードと比較し, そのユークリッド距離が特徴マップの勝者ノードとの距離より小さい場合, 入力ベクトルは記憶部に保存された勝者ノードのクラスに属する. そうでなければ, 特徴マップの勝者ノードとのユークリッド距離の上限 (閾値) を設定し, 入力ベクトルにクラス番号を付ける (学習ベクトル量子化 (LVQ) アルゴリズム⁵⁾).

• **step 5:** 行動学習. 行動マップのノードの入力は特徴マップによる分類した入力ベクトルのクラスの番号となる. 行動マップ上のノードの出力はロボットの行動番号とする. 命令に従う行動の獲得は強化学習アルゴリズムによって実

現され, その詳細は 2.3 節で述べる.

• **step 6:** 記憶と追加学習. 行動学習において, 十分な報酬が与えられた場合, その命令の学習が終了する (2.3 節を参照). このとき, 特徴マップの勝者ノードのベクトル値, 特徴クラス番号, および行動マップ上の行動番号を記憶部に保存し, 勝者ノードのベクトル値およびそれに対応した行動マップのラベルを初期値 (ランダム値) に戻す. よって, リフレッシュされたノードは, 再学習または追加学習に利用することができる. ただし, 再学習や追加学習の場合は **step 1** に戻る.

上記のアルゴリズムによって十分学習した特徴マップのノードと行動マップのノードが記憶部に保存されるため, そのクラスの入力データはつぎに提示されたとき, 記憶部を参照するだけで分類できる. また, 特徴マップの SOM のサイズはあらかじめ固定されているが, ノード値がリフレッシュできるため, Transient-SOM は永続的に入力データを分類することができる.

入力データの分類を行なう際に, 特徴空間の定義は分類の結果を大きく左右する. ここでは平行移動・回転・拡大処理によって正規化された画像に対して, 手首 (最長軸の端点) から輪郭線までの手領域を $18^\circ \sim 162^\circ$ の間で 1.8° ずつ反時計回りに分割し, その長さを特徴量 (80 次元) とする²⁾.

2.3 強化学習による行動マップの更新

強化学習システムにおいて, エージェントは異なる環境状態における確率的な行動方策に従って行動し, その結果, 環境から報酬または罰をもらうことによって, 行動価値関数が更新され, 行動方策もより報酬がもらえるように修正される⁷⁾. Transient-SOM では, 特徴マップの各ノードに付けたクラス番号は行動マップへの各ノードの入力で, それを環境の状態と見なす. また, 行動マップの出力はロボットが選択した行動の番号と見なすことができる. ある状態 s_t におけるロボットの行動 a_t^i は $\pi(a_t^i | s_t) = \frac{\exp(Q(s_t, a_t^i)/T)}{\sum_{j \in A} \exp(Q(s_t, a_t^j)/T)}$ という確率方策 (ボルツマン分布) に従って選択される (ソフトマックス行動選択規則). ここでは, T は温度パラメータ, $s_t = 1, 2, \dots, N \times M$. a_t^i は選択可能な行動, $i = 1, 2, \dots, K, K < N \times M$ である. $Q(s_t, a_t^i)$ は状態行動価値関数で, ランダムな正または負の整数の初期値をもち, $Q^\pi(s_{t+1}, a_{t+1}^i) = Q^\pi(s_t, a_t^i) + r$ によって更新される. ただし, r は状態 s_t で行動 a_t^i を行なったとき, 命令指示者によって与えられた報酬または罰を表わす (ここでは正または負の整数値とする). Q 値に正の閾値を設けることによって, 学習の終了条件が設定される.

2.4 感情マップ

感情マップ (Feeling map) は行動マップと同様に, $N \times M$ 個のノードをもつ 2 次元マップでそれぞれのノードは区間 $(-1, 1)$ の感情値 $Feel_i$ をもつ. 感情マップ上の各ノードは入力, 特徴マップ, 記憶層の各ノードと結合している. 初期値 $Feel_i = 0.0$, 褒められた場合, または, 叱られた場合, それぞれ, $Feel_i \leftarrow Feel_i \pm \alpha * C - \beta * SE_i$ により更新され

る。ただし、 C は褒められ (叱られ) 続けた回数, SE は入力ベクトルと勝者ベクトルとの距離 (2 乗誤差), パラメータ α と β は $-1 < \alpha < 1, 0 < \beta < 1$ である。

3. 実機実験

ソニー (株) のエンターテインメントロボット AIBO (ERS-7, 2003)⁸⁾ を用いて提案した手画像命令学習システムの有効性を検証した。実験は室内・蛍光灯照明の条件下で行なわれた。命令教示者は日本人女性で手首までの長袖を装着した。AIBO の頭センサーを触る (褒める場合) または背中センサー (後ろ) を触る (叱る場合) ことによって行動したロボットに報酬または罰を与えた。特徴マップ, 行動マップ, 感情マップのサイズはいずれも 5×5 とした。AIBO による入力画像 (YCrCb 表色系) のサイズは 160×208 pixels であった。近傍関数 $h_{ci}(t)$ の値は注目ノードから六角形のマップ⁵⁾ 上のノードとの距離によって, 1 単位離れるごとに, 0.5, 0.3, 0.1 の値を与えた。LVQ に用いられたユークリッド距離の閾値は 500.0 とした。学習回数を減らすため, $T = 0.0$ と見なして, 単純グリーディ行動選択を用いた⁷⁾。報酬または罰の値は 1 回の行動で 10.0 または -10.0 とした。学習終了条件は $Q > 50.0$ で, パラメータ α と β の値はそれぞれ 0.1, 0.0001 とした。

教示者の手の形状が「グー」, 「チョキ」, 「パー」, 「4 本指」の順で, それぞれに対応すべきロボットの行動は「伏せる」, 「座る」, 「立つ」, 「お辞儀する」の 4 種類の命令はすべて学習できた。特徴マップでの競合学習回数は 100 回, 行動マップでの強化学習回数は各命令 25 回計 100 回とした。Fig. 2 は学習に伴う各層の状態の変化を示している。左側の特徴マップでは各ノードが受けた入力ベクトルの値を示している (横軸は次元番号, 縦軸は肌色領域の長さ)。中間の行動マップの数値は前述のロボットの 4 つの行動番号を意味する。右側の感情マップの色の濃淡は感情値 (嬉しさ, 悲しさ) の大きさを表わしている。なお, 図中の実線の円が付いたノードは十分に学習を行ない, 記憶部に保存され, 点線の円が付いたノードはその値がリフレッシュされたことを表わす。AIBO の顔には赤, 白, 青のランプがあり, その点灯を制御することによって, 学習中のロボットの感情値を表情に表わすことができた。なお, 教示者 A が訓練した命令システムは, 利用者 B の手の命令も識別できることが確認された。カメラに手の形を提示する際の角度の限界は, カメラの光軸を中心に左右約 90° , 縦軸を中心に前後約 30° (パン), 横軸を中心に上下約 45° (チルト) であった。

4. ま と め

本論文では特徴マップ層・記憶層・行動マップ層・感情マップ層をもつ手画像命令学習システムを提案した。AIBO を用いた実機実験で提案システムの実用性を確認した。また, システムの中核である Transient-SOM は音声など手画像以外の命令学習システムへの応用も考えられる。

Fig. 2 Training results of Feature map (left), Action map (center), Feeling map (right)

謝辞 本研究の一部は日本学術振興会科学研究費補助金基盤研究 (C)(No.18500230) によるものである。

参 考 文 献

- 1) パートナーロボット資料集成, (株)NTS (2005)
- 2) N. Tanibata, N. Shimada and Y. Shirai: Extraction of Hand Features For Recognition of Sign Language Words, in Proc. Intern. Conf. on Vision Interface, 391/398 (2002)
- 3) J. Sherrah and S. Gong: CVonline: Skin Colour Analysis (2001) <http://homepages.inf.ed.ac.uk/rbf/CVonline/>
- 4) T. Kuremoto, T. Hano, K. Kobayashi and M. Obayashi: A hand instruction learning system using Transient-SOM, in Proc. of the 2nd Intern. Conf. on Natural Computation and the 3rd Intern. Conf. on Fuzzy Sys. and Knowledge Discovery, 403/414 (2006)
- 5) <http://www.cis.hut.fi/nnrc/refs/>
- 6) 道端, 三好, 増山: 自己組織化マップの自動追加学習システム, 日本ファジィ学会誌, 14-2, 147/156 (2002)
- 7) R.S. Sutton and A.G. Barto: Reinforcement Learning: An Instruction, The MIT Press, London (1998)
- 8) AIBO Official Site: <http://www.jp.aibo.com/aibostory/>