

## 動作の早期認識のための自己組織化写像構成法

川島 学<sup>†1</sup> 島田 敬士<sup>†1</sup> 谷口 倫一郎<sup>†1</sup>

本稿では、自己組織化写像を利用した動作の早期認識手法を提案する。動作の早期認識とは、入力された動作が何であるかをできるだけ早い段階で判断するものである。早期認識において認識する動作の初期部分が複数の種類の動作で共通である場合、動作初期の共通部分が入力されている間は認識結果の確定が不可能である。このため早期認識では動作初期の共通部分が入力されている間は結果を確定せず、共通部分の終了時点で結果を確定するようにしなければならない。提案手法では各動作の尤もらしさを確率として出力することでこの問題を解決する。結果を確率として出力することで、どの程度認識結果が信頼できるかを判断することができ、出力確率を利用して結果を確定することも可能である。実験では提案手法の有効性を確認した。

### Early Gesture Recognition using Self-Organizing Map

MANABU KAWASHIMA,<sup>†1</sup> ATSUSHI SHIMADA<sup>†1</sup>  
and RIN-ICHIRO TANIGUCHI<sup>†1</sup>

We propose a new approach for early gesture recognition using Self-Organizing Map. Early gesture recognition is a method to recognize sequential patterns at their beginning parts. If some gestures have a common gesture element, which is partial gesture included in the gesture, it is not easy to distinguish their gestures. Therefore, an early recognition result should be output after such a common gesture element. Our approach handles this problem by outputting plausibility of each gesture. This enables us to evaluate which gesture likely to be observed. We confirmed the performance of our approach in some experiments.

<sup>†1</sup> 九州大学  
Kyushu University

### 1. はじめに

マンマシンインタラクションや、アバタと呼ばれるキャラクターを用いた人間同士のコミュニケーションでは、人間がどのような動作をしているかという情報を取得することが重要である。人間の動作の認識は、これらのシステムにおいて重要な役割を担うものであり、盛んに研究が行われている。従来の多くの動作認識は、動作の入力が完了してから認識を行うものであったが、動作の入力が完了してから認識では、上記のシステムのような実時間のコミュニケーションにおいては違和感を感じてしまう。例えばマンマシンインタラクションでは、人間が行った動作を機械が認識し、認識した動作に対応する行動をとる必要があるが、動作が完了してから認識処理を開始すると、人間の動作終了から機械の行動開始までに遅れが生じてしまい、違和感を感じてしまう。この問題を解決する手法として、動作入力初期段階で認識結果を確定する早期認識という方法がある。早期認識が可能であれば、動作入力の完了前にシステムが対応を取れるため、上記のシステムにおける遅れを補償でき違和感のないコミュニケーションが可能となる。よって本研究では、早期認識の手法を確立することを目的とする。

早期認識の手法として、連続 DP による方法<sup>1),2)</sup>、Boosting 的学習法<sup>3)</sup> が提案されている。連続 DP による動作認識は、動作の時間長の変動を考慮した類似度を計算する手法であり、これを早期認識に対応させ、この性質が継承された早期認識手法を実現している。また、ジェスチャネットワークと呼ばれる部分動作間をつなぐネットワークによって共通部分動作の関係を記述することで、早期認識の結果を確定している。しかし、このネットワークの作成は自動化されておらず、実験では手動でネットワーク作成を行っている。ジェスチャネットワークは対象とする動作の集合に応じて動的に構成されるものであるため、手動でのネットワークの作成は非常に手間のかかる作業である。Boosting 的学習法では、各時刻に識別器(弱識別器)を用意し、それらを統合して早期認識結果を得ている。早期認識の結果を確定するには、文献<sup>4)</sup>の手法が応用可能であるとしている。しかし、この手法では時間長変動問題が考慮されていない。時間長変動問題とは、動作の長さがクラス間、また個々のパターン間で異なるという問題である。このような異なる時間長の動作入力に対しても、正しく認識結果を出力する必要がある。

本研究では動作の早期認識手法を提案するにあたって、島田らの提案した動作認識の手法である自己組織化写像(Self-Organizing Map: SOM)のスパースコードを用いた動作認識手法<sup>5)</sup>を利用する。SOMは人工神経回路網モデルのひとつであり、動作の構成要素である

時系列姿勢を学習する．SOM に姿勢を入力すれば，その姿勢に最も類似した姿勢を記憶しているニューロン（以下では，最大活性化ニューロンと呼ぶ）が得られる．従って，動作を構成する時系列姿勢を順にすべて入力すると，その時系列姿勢に対して複数のニューロンが最大活性化ニューロンとして選ばれることになる．島田らは，マップ全体の最大活性化ニューロンと活性化しなかったニューロンを利用して，動作をマップ全体のニューロンの活性化パターン（スパースコード）として捉えることで動作認識を行っている．スパースコードを利用することで時間長が変動する動作を効率よく認識することが可能である．本研究ではこれを早期認識に対応させ，動作の早期認識における時間長変動問題に対処する早期認識手法を提案する．

早期認識を行う複数の動作に類似姿勢（共通部分動作）が含まれている場合，このような姿勢については，同じあるいは近傍のニューロンが最大活性化ニューロンとして選択される．このようなニューロンからは，どの動作が入力されたかを確定するのは難しい．一方，ある動作に含まれる姿勢にしか最大活性化ニューロンになり得ないニューロンが分かれば，そのニューロンの活性化状況を調べることで，動作認識結果を確定することが可能である．しかし，SOM の性質上類似姿勢は近傍のニューロンに写像されるため，このような動作固有のニューロンの周辺も，ある程度動作固有のニューロンであるということになる．よって最大活性化したニューロンが動作固有のニューロンであるかは確定的ではなく，判断をどのように行うかが問題となる．

そこで本研究では認識結果を確定するための指標として，各動作の尤もらしさを確率として出力する 2 つの手法を提案する．本手法の利用者は出力された確率を利用した認識結果の確定方法を決定することができる．実験では閾値を設定することにより認識結果を確定している．提案手法 1 は，出力する確率にベイズ推定の最大事後確率を用いる手法である．提案手法 2 では，動作の姿勢を SOM に入力し活性化ニューロンを求め，そのニューロンを活性化させ得る動作がどの動作であるかを確率として出力する．以下では，提案手法の要素となる技術と提案手法について述べ，実験結果を報告する．

## 2. 自己組織化写像による動作認識

### 2.1 自己組織化写像

自己組織化写像 (SOM) は，人工神経回路網モデルの一つで，中間層の無い 2 層型の教師なし競合近傍学習モデルである<sup>6)</sup>．競合近傍学習とは，学習においてニューロンの結合係数を更新する際に，近傍のニューロンも更新する学習法である．SOM の学習アルゴリズム

をまとめると次のようになる．

Step 1 マップ上すべてのニューロン  $u$  の結合係数ベクトル  $W_u$  にランダムな値を設定する．

Step 2 入力ベクトル  $I$  に対する最大活性化ニューロン  $c$  を式 (1) によって決定する．

$$c = \underset{u}{\operatorname{argmin}}(\|I - W_u\|) \quad (1)$$

Step 3 すべてのニューロン  $u$  に対して式 (2) を適用し，結合係数ベクトル  $W_u$  を更新する． $h_{c,u}$  は更新する際の係数， $r_u$  はニューロン  $u$  のマップ上の位置である． $\alpha(t)$  は学習率， $\sigma(t)$  は近傍範囲を制御するパラメータである (式 (4)，(5))． $t$  は離散時間， $T$  は学習回数の上限である． $\alpha_0$ ， $\sigma_0$  はそれぞれ  $\alpha$ ， $\sigma$  の初期値である．

$$W_u(t+1) = W_u(t) + h_{c,u}(I - W_u(t)) \quad (2)$$

$$h_{c,u} = \alpha(t) \exp\left(-\frac{\|r_c - r_u\|^2}{2\sigma^2(t)}\right) \quad (3)$$

$$\alpha(t) = \alpha_0\left(1 - \frac{t}{T}\right) \quad (4)$$

$$\sigma(t) = \sigma_0\left(1 - \frac{t}{T}\right) \quad (5)$$

Step 4  $t < T$  であれば，Step2 に戻る．

### 2.2 スパースコーディングによる動作の表現

神経科学の分野では，脳は複数のニューロンの活性化パターンによって対象を認識していると考えられている．この考え方をスパースコーディングと呼び，生成された符号をスパースコードという<sup>7)</sup>．島田らは SOM にスパースコーディングの考え方を適用し，動作認識を行っている．以下では，スパースコーディングによる動作の表現方法について述べる．

#### 2.2.1 姿勢情報の学習

はじめに，動作の基本要素である姿勢情報を SOM に学習させる．入力ベクトル  $I$  は，姿勢情報を表す特徴ベクトルである．認識したい動作クラス  $m(1, \dots, m, \dots, M)$  を構成するすべての姿勢情報を SOM に学習させることで，類似した姿勢情報をもつニューロンが近くに存在するマップができる．

#### 2.2.2 活性化状況の記録

姿勢情報の学習後，各動作クラスを SOM に入力した際のニューロン活性化状況を記録する．ある動作を構成するフレーム数を  $T$  とすると，フレーム  $t(1, \dots, t, \dots, T)$  の姿勢は  $I(t)$  と表現できる． $I(t)$  に対する最大活性化ニューロンを  $c(t)$  とすると，動作を構成する全フ

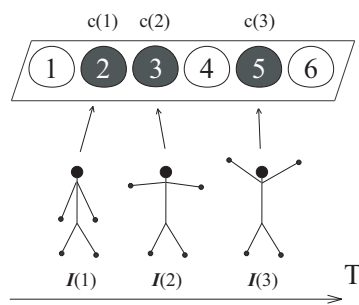


図 1 動作に対する活性化ニューロン

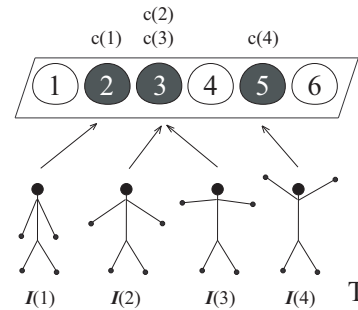


図 2 スパースコーディングによる時間長吸収の様子

フレーム  $t$  に対して  $c(t)$  を求めることができる．ニューロン  $u(1, \dots, u, \dots, N)$  が動作  $m$  を構成する時系列姿勢に対して活性化したかどうかをニューロン  $u$  の活性化状態  $s_{u,m}$  により表すことにする．全ニューロンにおける  $s_{u,m}$  の値をマップ上にパターンとして表したものが入力動作を表すパターン，すなわちスパースコードとなる．

これを図 1 を用いて説明する．この図では，上段に SOM のニューロン  $u(1, \dots, u, \dots, 6)$ ，下段には入力する動作を示している．この例は両手をあげる 3 フレームの動作である．図 1 では，動作を構成する各フレームの姿勢に対して最大活性化したニューロンに色付けをし，どの姿勢に対しても最大活性化ニューロンとならなかったニューロンには色付けを行っていない．このニューロンの色付けのパターンがスパースコードである．

これを一般化すると次のようになる．まず，SOM の全ニューロン  $u$  の活性化状態  $s_{u,m}$  は式 (6) のように表現できる．

$$s_{u,m} = \begin{cases} 1, & \text{if } u = c(t) | t \in T, m \in M \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

動作  $m$  を表す時系列パターン

$I = (I(1), \dots, I(t), \dots, I(T))$  に対するスパースコード  $S_m$  は，全ニューロンの活性化状態  $s_{u,m}$  をベクトルの要素として並べたものとなる (式 (7)) ．

$$S_m = (s_{1,m}, \dots, s_{u,m}, \dots, s_{N,m}) \quad (7)$$

ところで，動作の各姿勢に対する最大活性化ニューロンが重複することがあるが，その場合も  $s_{u,m} = 1$  として扱う．この時の様子を表したものが図 2 である．この例の動作は図 1

より 1 フレーム長い動作であるが，フレーム 2, 3 に対して同じニューロンが最大活性化することで時間長の違いを吸収している．これにより，時間長の異なる同じクラスの動作であっても，類似したスパースコードを生成することができ，同時に時間長の変動を吸収することが可能である．

### 2.2.3 スパースコードの平滑化

SOM の性質上，近傍のニューロンには類似した結合係数ベクトルが記憶されるため，同じ動作クラスであっても最大活性化ニューロンの位置に若干のずれが生じることがある．そこで，式 (7) に示したスパースコードに，ガウシアンフィルタによる平滑化処理を加える．これにより，スパースコード内の活性化ニューロンの位置にずれが生じてても，式 (1) の類似度計算を適切に行うことができる．ニューロン  $u$  の位置を  $r_u$  としたとき，この処理を式で表わすと式 (8) となる． $\sigma^2$  はガウシアンフィルタの分散である．

$$s_{u,m} = \sum_v \left\{ s_{v,m} \exp \left( -\frac{\|r_v - r_u\|^2}{2\sigma^2} \right) \right\} \quad (8)$$

## 3. 動作の早期認識

本章では，スパースコードを利用し，早期認識の結果を確率として出力する次の 2 つの早期認識の手法について述べる．

手法 1 最大事後確率を利用した早期認識

手法 2 ニューロンの活性値を利用した早期認識

手法 1 は，ベイズ推定によって前のフレームの確率を更新する手法である．これにより，時系列を考慮しないと結果を確定できないような入力に対しても正しく認識が行えることが期待される．これは例えば，ある動作である確率が比較的高いが他の動作の確率との差が 1 フレームでは認識結果を確定できない程度しかないような入力が挙げられる．

手法 2 は活性値を用いる手法である．活性値とは，ニューロン  $u$  の動作  $m$  に対する活性化状態  $s_{u,m}$  のことである．手法 1 では活性値を用いて前のフレームの確率を更新するが，手法 2 では前のフレームの確率を用いず当該フレームの活性値のみで出力確率の計算を行うため，より単純な計算で結果の出力が可能である．

### 3.1 最大事後確率を利用した早期認識

動作クラス  $m$  が観測される事象を  $A_m$ ，ニューロン  $u$  が最大活性化する事象を  $X_u$  とする．フレームを  $t$  とすると，求めたい確率は，ニューロン  $u$  が活性化した時，観測された

動作がある動作クラス  $m$  であるという確率  $P_t(A_m|X_u)$  である。この確率をベイズ推定で表すと式 (9) となる。提案手法 1 ではこの式で計算した値を出力する。結果を確定するときには、例えばこの確率  $P_t(A_m|X_u)$  が閾値  $th_1$  の値を越えた時、動作クラス  $m$  を認識結果として出力する。

$$P_t(A_m|X_u) = \frac{P(X_u|A_m)P_t(A_m)}{\sum_m P(X_u|A_m)P_t(A_m)} \quad (9)$$

式 (9) の  $P(X_u|A_m)$  は尤度を表す式である。すなわち、動作クラス  $m$  が観測されたとき、ニューロン  $u$  が活性化する確率を表している。これを式で表すと、式 (10) となる。

$$P(X_u|A_m) = \frac{s_{u,m}}{\sum_u s_{u,m}} \quad (10)$$

$s_{u,m}$  は、動作クラス  $m$  に対するスパースコードのニューロン  $u$  の状態を表している。式 (6) は、動作クラス  $m$  に属する姿勢のうち、ニューロン  $u$  が活性化する姿勢がひとつでもあれば  $s_{u,m} = 1$  となることを示している。すなわち、動作クラス  $m$  の学習データのいずれかの姿勢に対して、ニューロン  $u$  が活性化したかどうかを表している。この値を次のように解釈することで、尤度の計算に利用することができる。ニューロンの状態  $s_{u,m}$  は、動作クラス  $m$  が観測されたときにニューロン  $u$  が活性化すれば 1、しなければ 0 となる。これは、動作クラス  $m$  が観測されたという条件のもとでニューロン  $u$  が活性化するという事象  $X_u|A_m$  をそのまま表していると考えることができる。よって、この事象を条件付き確率  $P(X_u|A_m)$  として表す場合、式 (10) に示すように、ニューロン全体に対する  $s_{u,m}$  の割合を計算すれば良いことになる。これが尤度  $P(X_u|A_m)$  である。

式 (9) における事前確率  $P_t(A_m)$  は、動作クラス  $m(1, \dots, m, \dots, M)$  がどれくらいの確率で観測されるかという事前知識である。式 (11) は、式 (9) における事前確率を表す式であり、入力時の活性化ニューロンの数が多い動作クラスほど事前確率が高くなるようにしている。

$$P_t(A_m) = \begin{cases} \frac{\sum_u s_{u,m}}{\sum_m \sum_u s_{u,m}} & \text{if } t = 1 \\ \frac{P_{t-1}(A_m|X_{c(t-1)}) \sum_u s_{u,m}}{\sum_m \left\{ P_{t-1}(A_m|X_{c(t-1)}) \sum_u s_{u,m} \right\}} & \text{otherwise} \end{cases} \quad (11)$$

自然に考えれば、どの動作クラスが抽出されるかはわからないため、事前確率はどれも同じである。しかし、動作クラス間には観測のされやすさに違いがあると考えられる。まずあるひとつの入力姿勢に注目すると、認識させる全クラスの動作を構成する全姿勢の中からひとつの姿勢が抽出され、観測されると考えることができる。よって、フレーム数が多い動作クラスほど抽出される確率が高いと考えることができる。これを SOM のマップ上に写像して考えると、動作入力時に最大活性化したニューロンの数が多いほど観測されやすいということになる。ある動作クラス  $m$  を入力した時に最大活性化したニューロンの数は、ニューロン状態  $s_{u,m}$  を利用して、 $\sum_u s_{u,m}$  と表すことができる。この値が各動作クラスの観測されやすさとなるため、これを動作クラス全体で 1 になるように正規化すると式 (11) に示した  $t = 1$  の事前確率  $P_1(A_m)$  となる。 $t = 2$  以降の観測についても同様に、各動作クラスに対する活性化ニューロンの数によって事前確率を補正している。フレームの長さではなく、活性化ニューロンの数で事前確率を設定することにより、動作の時間長の違いによらない事前確率を設定することができる。

以下に、早期認識手法としての全体のアルゴリズムを示す。フレーム  $t$  の初期値は 1 である。Step 7 の閾値処理は、本研究で行った実験で明確な認識率を出すために便宜的に加えた処理である。確率として結果を出力するのであれば Step 3 で計算した事後確率  $P_t(A_m|X_u)$  を出力し、Step 2 に戻る。

Step 1 SOM に全クラスの動作を構成する全姿勢を学習させる。SOM のニューロンが持つ結合係数ベクトルは、姿勢の特徴ベクトルである。

Step 2 各動作クラス  $m(1, \dots, m, \dots, M)$  に対するスパースコード  $S_m$  を式 (6), (7) により求める。

Step 3 Step 2 で求めたスパースコードを、式 (8) のガウシアンフィルタにより平滑化する。

Step 4 SOM の全ニューロン  $u$ 、全動作クラス  $m$  について、尤度  $P(X_u|A_m)$ 、初期フレームの事前確率  $P_1(A_m)$  をそれぞれ式 (10)、式 (11) により求める。

Step 5 入力動作のフレーム  $t$  の姿勢  $I(t)$  に対する SOM の活性化ニューロン  $c(t)$  を式 (1) により求める。

Step 6 すべての動作クラス  $m$  について、事後確率  $P_t(A_m|X_{c(t)})$  を式 (12) で計算する。

$$P_t(A_m|X_{c(t)}) = \frac{P(X_{c(t)}|A_m)P_t(A_m)}{\sum_{A_m} P(X_{c(t)}|A_m)P_t(A_m)} \quad (12)$$

Step 7 閾値  $th_1(\frac{1}{M} \leq th_1 < 1)$  に対して,  $P_t(A_m|X_{c(t)}) > th_1$  となる事後確率の値があれば, その動作クラス  $m$  を認識結果として出力する. なければ式 (11) によって次フレームの事前確率の計算を行った後, Step 2 に戻る.

### 3.2 ニューロンの活性化値を利用した早期認識

この手法では動作を構成するひとつの姿勢を入力し, ニューロン  $u$  が最大活性化した時, 入力された動作が動作クラス  $m$  である確率  $P(A_m|X_u)$  が求められる. 提案手法 1 では前のフレームの事後確率  $P_{t-1}(A_m|X_u)$  を用いて, フレーム  $t$  の確率  $P_t(A_m|X_u)$  を計算する必要があったが, この手法では前のフレームの情報を使わずに, フレーム  $t$  で最大活性化したニューロン  $c(t)$  のみを求めることで, 出力確率を求めることができる. すなわち, 各ニューロンごとに事前に出力確率の計算ができるということである. よって認識過程では, SOM の活性化ニューロンの計算のみで結果を出力することができる. この手法によって出力される確率を式 (13) で表わす.

$$P(A_m|X_u) = \frac{s_{u,m}}{\sum_m s_{u,m}} \quad (13)$$

動作クラス  $m$  に対するニューロン  $u$  の状態  $s_{u,m}$  は, 動作の入力に対してニューロン  $u$  が活性化したとき 1 になる. ニューロン  $u$  に注目すると,  $s_{u,m} = 1$  であるということは, ニューロン  $u$  が活性化したとき, 動作  $m$  が入力された可能性があるということを表している. この可能性を確率として表現するために  $s_{u,m}$  を動作全体で 1 となるように正規化すると, 式 (13) となる. すなわち, スパースコードのニューロン  $u$  における活性化状態  $s_{u,m}$  は, ニューロン  $u$  が活性化した時, 入力された動作の動作クラスが  $m$  であるという事象  $A_m|X_u$  をそのまま表していると解釈しているということである. この式は入力フレーム  $t$  を含まないので, 提案手法 1 とは異なり, 動作の入力前に  $P(A_m|X_u)$  の値を全ニューロン  $u$  に対して求めることができる.

以下に, 早期認識手法としての全体のアルゴリズムを示す. フレーム  $t$  の初期値は 1 である. Step 7 の閾値処理は, 提案手法 1 のアルゴリズムと同様に, 本研究で行った実験で明確な認識率を出すために便宜的に加えた処理である. 確率として結果を出力するのであれば Step 6 で求めた確率  $P(A_m|X_{c(t)})$  を出力し, Step 5 に戻る.

動作番号	フレーム数	内容
動作 1	9 ~ 100	両手を広げて肩に乗せる
動作 2	7 ~ 90	両手を前方に突き出す
動作 3	7 ~ 95	背負い投げの姿勢をする
動作 4	12 ~ 127	両手を体の前でぐるぐる回す
動作 5	19 ~ 259	両手を交互に上げ下げする
動作 6	18 ~ 320	右手をひねる

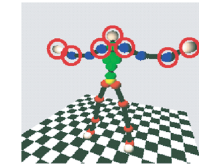


図 3 実験に利用した特徴点

Step 1 SOM に学習データとなる動作を構成する全姿勢を学習させる. SOM のニューロンが持つ結合係数ベクトルは, 姿勢の特徴ベクトルである.

Step 2 各動作クラス  $m(1, \dots, m, \dots, M)$  に対するスパースコード  $S_m$  を式 (6), (7) により求める.

Step 3 Step 2 で求めたスパースコードを, 式 (8) のガウシアンフィルタにより平滑化する.

Step 4 SOM の全ニューロン  $u$ , 全動作クラス  $m$  について, 確率  $P(A_m|X_u)$  を式 (13) により求める.

Step 5 入力動作のフレーム  $t$  の姿勢  $I(t)$  に対する SOM の活性化ニューロン  $c(t)$  を式 (1) により求める.

Step 6 すべての動作クラス  $m$  について, 確率  $P(A_m|X_{c(t)})$  を求める.

Step 7 閾値  $th_2(\frac{1}{M} \leq th_2 < 1)$  に対して,  $P(A_m|X_{c(t)}) > th_2$  となる確率の値があれば, その動作クラス  $m$  を認識結果として出力する. なければ Step 2 に戻る.

## 4. 実験と考察

### 4.1 実験の設定

入力する動作は, カーネギーメロン大学の Graphics Lab が公開しているモーションキャプチャデータベース (<http://mocap.cs.cmu.edu/>) から取得したものを利用している. このデータは, 体の 25 か所にマークを取り付けて計測されたものである. 各マークの 3 次元位置を, 姿勢を表す特徴ベクトルとして利用した. 今回, 実験で用いる動作クラス数は 6 とした. 各動作の内容を表 1 に示す. 表からわかるように, 各動作の動作長は動作クラス間, また同じクラス内でもかなり変動している.

実験は学習データが各動作クラス 20 個, テストデータは各動作クラス 10 個で行った. 本

ニューロン数 $N$	50 × 50
次元数 $d$	21
学習回数 $T$	姿勢数 × 100
学習率 $\alpha_0$	0.9
標準偏差 $\sigma_0$	9.26

ガウシアンフィルタの分散 $\sigma^2$	4.0
提案手法 1 の閾値 $th_1$	95%
提案手法 2 の閾値 $th_2$	80%

動作番号	提案手法 1	提案手法 2
動作 1	100	100
動作 2	20	100
動作 3	100	90
動作 4	100	100
動作 5	90	100
動作 6	70	80
平均	80	95

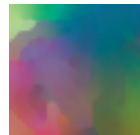


図 4 SOM の学習結果

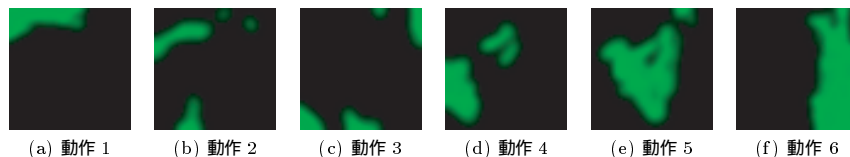


図 5 各動作クラスに対するスパースコード

実験での入力姿勢は、図 3 に示した 7 つの特徴点を利用した。これは実験で利用した動作がすべて上半身の動きで表現できたからである。

実験で利用した SOM の構成を表 2 に、その他のパラメータの設定を表 3 に示す。ガウシアンフィルタの分散とは、式 (8) における  $\sigma^2$  のことである。

#### 4.2 SOM の学習結果とスパースコード

まず SOM の学習後のマップを図 4 に示す。結果は学習データの全フレームを主成分分析し、ニューロンの結合係数ベクトルの第 3 主成分までを RGB 値として色づけしたものを示している。学習の結果、類似した色同士が近くに写像されているため、マップ上に類似特徴が記憶されていることがわかる。

各動作クラスに対する SOM のスパースコードを図 5 に示す。この図では、動作クラス  $m$  に対するニューロン  $u$  の活性化状態  $s_{u,m}$  の値を緑色の成分で示している。結果から、各動作クラスのスパースコードが異なるものになっていることがわかる。また、比較的長い動

作である動作 5、動作 6 は、他の動作に比べて構成する姿勢の数も多いため、活性化ニューロンが多くなっていることもわかる。

#### 4.3 各手法の早期認識結果

図 6 に提案手法 1、図 7 に提案手法 2 の各動作クラスの入力に対する出力確率を示す。横軸がフレーム数、縦軸が確率を表す折れ線グラフとなっている。グラフは、動作クラス内で最もフレーム数が長い動作について示している。また、各動作は長さの異なるものであるが、動作初期の変動を見やすくするため、初期 50 フレームの結果のみを示している。

提案手法 1 では、動作 1、3、4 では、かなり初期のフレームで早期認識が達成されているが、動作 5、動作 6 は間違った動作の確率が大きくなったあと、正しい動作に認識されている。動作 2 に関しては、初期段階で動作 1 に誤認識されたあと、正しい動作である動作 2 に認識されるという結果になった。

一方、提案手法 2 でも、動作 1 と動作 4 については、ほぼ初期の段階で正しく認識されている。動作 3 は、初期段階では動作 3 と動作 6 の値が比較的大きい値を示し、それ以降では正しく認識されている。また、動作 3 の値は初期段階でも 80% を超えているが、動作 6 は超えていない。動作 5 と 6 は、初期段階はどの動作も 80% より低い確率の値を示し、その後正しい動作に認識されている。動作 2 に関しては、初期段階では動作 1 の値が比較的高いが 80% より低く、その後正しく認識されるという結果になった。

表 4 は、各動作クラス 10 個のテストパターンに対して、閾値処理での早期認識を行った結果である。認識結果は提案手法 2 の方が全体的に高く、動作 2 の認識率は特に改善されている。

#### 4.4 考 察

表 1 に示したように実験に用いた動作は動作クラス間、動作クラス内で長さの異なるものであるが、提案手法は正しく早期認識を行うことができている。よって、提案手法は図 2



に示した動作長の違いの吸収を早期認識においても行えているといえる。

提案手法1の実験の結果、誤認識した動作について調査した。その結果、各動作に共通の姿勢である初期姿勢の部分で認識を誤っていることがわかった。この原因は、次のように考えられる。各動作に共通の姿勢が入力された時、大きくはないが各動作間で尤度の差が生まれる。その尤度の差は1フレームのみでは大きな事後確率の変化を起ささないが、その姿勢が続いてもその間ずっとベイズ改訂は続けられる。このため、確率の差が大きなものとなってしまい、誤認識が生じてしまう。つまり、各動作に共通の姿勢が長く続くと認識結果を誤ってしまう可能性があるということである。共通の姿勢が入力されたときはベイズ推定によってどの動作の確率も極端には大きくなるような結果を期待していたが、このような問題が起きてしまった。

一方、提案手法2の実験結果は、提案手法1よりも良好な結果が得られた。特に大きな改善が見られた動作2については、初期段階で動作1の確率が大きくなっているが、のちに確率が大きくなる動作2の確率より値が小さい。これを分離できる閾値を設定できれば、今回の実験のように正しい結果を得ることが可能である。最後まで早期認識の結果が確定しないような動作もあったが、結果に一定の信頼性を認められないなら出力しないという考えからいえば望ましい結果であるともいえる。また提案手法2は、前のフレームの情報を用いないので、事前にニューロンごとの確率が計算でき、ベイズ推定の計算も行う必要がないため、より単純で計算量も少ない。

## 5. おわりに

本研究では、自己組織化写像のスパースコードを利用した動作の早期認識の手法を2つ提案した。まず、SOMには動作を構成する姿勢を学習させ、動作をスパースコードで表現した。そして、提案手法1では、各フレームでの認識結果の出力にベイズ改訂の枠組みを利用して、早期認識を行った。また、提案手法2では、動作の姿勢をSOMに入力し活性化ニューロンを求め、そのニューロンの各動作クラスに対する活性化状態を確率として出力した。

実験では、閾値処理による早期認識を行った。認識の結果、提案手法1では各動作クラスに共通な部分の動作において確率出力の利点が発揮されず、間違った結果を出力してしまうことがあった。これに対し提案手法2では、各動作クラスに共通な姿勢の入力においてどの動作でも低い確率を出力することができた。認識率としては、提案手法1は全体で80%、提案手法2では95%となった。この結果により、提案手法1よりも、提案手法2の方が良

好な結果が得られることがわかった。さらに提案手法2は認識処理の計算も単純であるため、提案手法1よりも有用な手法であると考えられる。

今回行った提案手法2の実験では、閾値による早期認識結果の確定でも良好な結果が得られたが、この確定則の改善を行えば、さらに良い認識結果が得られる可能性がある。よって、今後の課題としては、閾値による手法以外の早期認識結果確定則の考案があげられる。他には、実時間システムに組み込むための動作予測手法の確立、実際のシステムへの組み込み等があげられる。

## 参 考 文 献

- 1) 内田誠一, 森明慧, 倉爪亮, 谷口倫一郎, 長谷川勉, 迫江博昭: 動作の早期認識およびその予測への応用に関する検討, 電子情報通信学会技術研究報告, PRMU, パターン認識・メディア理解, Vol.104, pp.pp.7-12 (2004).
- 2) 森明慧, 内田誠一, 倉爪亮, 谷口倫一郎, 長谷川勉, 迫江博昭: ジェスチャの早期認識・予測ならびにそれらの高精度化のためのネットワークモデルに関する検討, 画像の認識・理解シンポジウム (MIRU2005), pp.pp.1058-1065 (2005).
- 3) 天本一馬, 内田誠一: 時系列パターンの早期認識のための Boosting 的学習法, 画像の認識・理解シンポジウム (MIRU2008), pp.pp.1430-1435 (2008).
- 4) Sochman, J. and Matas, J.: WaldBoost - Learning for Time Constrained Sequential Detection, *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, Vol.2, pp.pp.150-156 (2005).
- 5) 島田敬士, 谷口倫一郎: 階層型自己組織化写像のスパースコードを利用した時系列動作パターンの認識, 画像の認識・理解シンポジウム (MIRU2008), pp.pp.394-399 (2008).
- 6) Kohonen, T.: *Self-Organizing Maps* (1995).
- 7) Olshausen, B.A. and Field, D.J.: Sparse coding of sensory inputs, *Current Opinion in Neurobiology*, Vol.14, pp.pp.481-487 (2004).

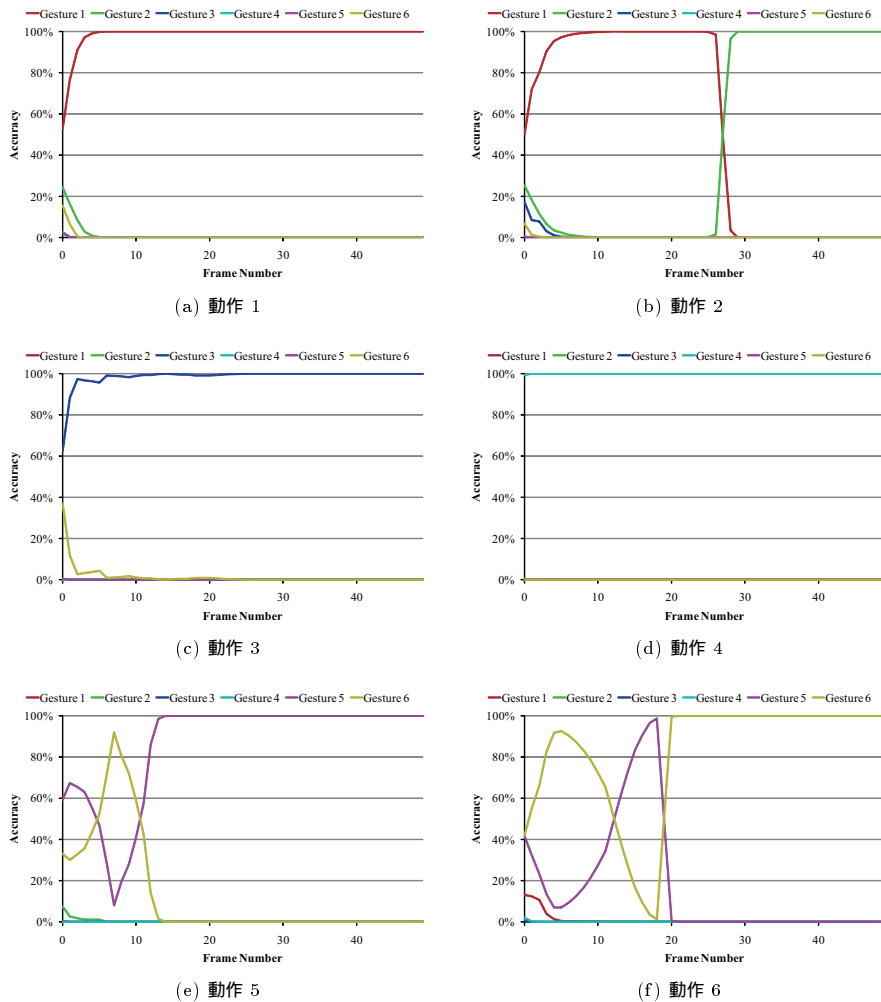


図 6 提案手法 1 の出力確率

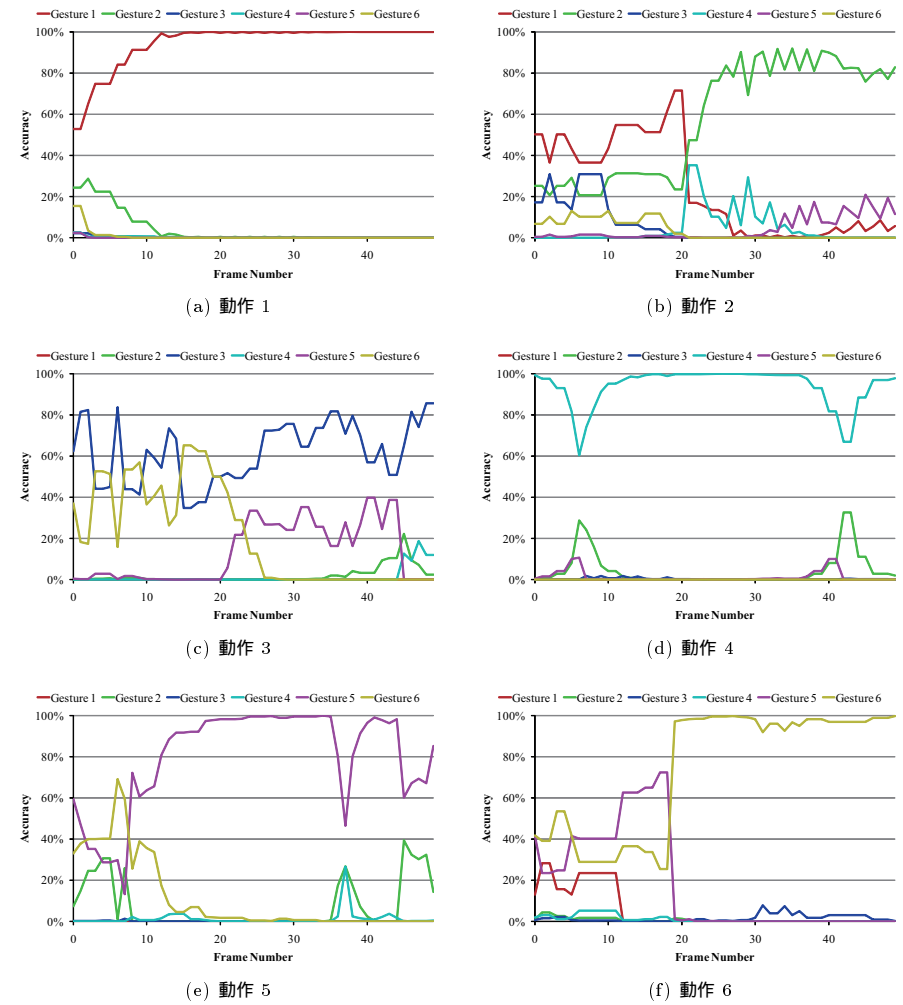


図 7 提案手法 2 の出力確率