

# 顔パターンの階層的位相変化に基づく表情特徴空間の形成

石井 雅樹<sup>†</sup> 佐藤 和人<sup>†</sup> 間所 洋和<sup>†</sup> 西田 眞<sup>††</sup>

<sup>†</sup> 秋田県産業技術総合研究センター 〒010-1623 秋田市新屋町字砂奴寄 4-21

<sup>††</sup> 秋田大学工学資源学部 〒010-8502 秋田市手形学園町 1-1

E-mail: †ishii@rdc.pref.akita.jp

あらまし 本稿では、教師無し学習の SOM ( Self-Organizing Map ) と教師あり学習の CPN ( Counter Propagation Networks ) を併用した表情特徴空間の形成手法を提案する。はじめに、表情の表出プロセスで発生する複雑な顔パターンの位相変化を狭い写像空間の SOM を用いて階層的に学習し、各表情を特徴付ける顔パターンの基準ベクトルを生成する。次に、各表情の基準ベクトルを CPN の入力とし、表情の位相変化に基づいた顔パターンのカテゴリマップを形成した後、無表情と基本 6 表情を基に各カテゴリに心理学的な意味付けを行う。なお、本稿ではこのカテゴリマップを表情特徴空間と定義する。被験者 6 名を対象とした実験結果より、提案手法は、顔画像上に表れる表情位相特性に基づき、自己組織的に被験者固有の表情特徴空間を形成できることを確認した。

キーワード 顔画像処理, 表情解析, 自己組織化マップ, カウンタープロパゲーションネットワーク

## Formation of Expression Feature Spaces based on a Hierarchical Topological Change of Facial Images

Masaki ISHII<sup>†</sup>, Kazuhito SATO<sup>†</sup>, Hirokazu MADOKORO<sup>†</sup>, and Makoto NISHIDA<sup>††</sup>

<sup>†</sup> Akita Prefectural R&D Center 4-21 Sanuki, Araya, Akita-City 010-1623 Japan

<sup>††</sup> Faculty of Engineering and Resource Science, Akita University 1-1 Tegata Gakuen, Akita-City, Akita 010-8502 Japan

E-mail: †ishii@rdc.pref.akita.jp

**Abstract** In this paper, we propose a formation method of a expression feature space with a combination of supervised and unsupervised learning. First, hierarchical SOM having a small mapping space learns topological changes of complex face patterns and generates a reference vector of face patterns to characterize each expression. Next, CPN learns reference vectors of each expression and forms a category map based on topological changes of face patterns. Finally, the category map is created emotional meaning based on neutral and six basic facial expressions (anger, sadness, disgust, happiness, surprise, and fear). We defined this category map as the expression feature space. The experimental results for six subjects indicated that the expression feature space defined individual expression features and the proposed method was effectively for forming feature space of individual facial expression.

**Key words** Facial image processing, Facial expression analysis, SOM, CPN

### 1. ま え が き

近年、情報機器の機能の高度化により、知的なマンマシンインタフェースの重要性が高まっており、人間の顔認識に関する研究が盛んに行われている [1] ~ [5]。中でも、Human Computer Interaction の観点から、マンマシン間の感情のコミュニケーションを目的とした顔表情の認識に関する研究が関心を集めている [1], [3], [6] ~ [15]。

代表的な表情分析モデルとして Ekman らによって提案され

た FACS ( Facial Action Coding System ) が知られている [16]。Ekman らによれば、人間の感情は、怒り、悲しみ、嫌悪、喜び、驚き、恐れ の 6 種類 ( 以下、6 基本表情と記す ) に分類されるとされており、FACS では、表情に関する眉、目、口などの顔部位の動き特徴が AU ( Action Unit ) として定性的に定義されている [16], [17]。これまでの多くの表情認識に関する研究では、この 6 基本表情に基づいたカテゴリ分類を試みている。しかしながら、顔および顔部位の形状に関する個人差、表情表出時の顔部位の動きの多様性などといった顔が有する特性から、

次のような課題が指摘されている [3] .

第一に、表情認識モデルの一般化に関する課題が挙げられる。多くの従来手法では、顔面上に設定した特徴点の移動量や任意領域内の表情筋の動ベクトルなどを特徴量とし、それらを AU と対応付けることにより、表情認識モデルの一般化を試みている (例えば [13])。FACS の AU は、曖昧で定性的な言語で記述されているため、特徴点および領域を顔面上のどの箇所に設定するかについては客観的な基準が存在せず、研究者の経験により定義されている。しかしながら、前述のように、表情を決定付ける特徴は被験者ごとに異なると考えられることから、全ての被験者に対して定量的な特徴と定性的な AU との関連付けを行い、普遍的な特徴を導き出すためには、膨大な労力が必要と考えられる。また、全ての被験者に対応した表情認識モデルの一般化により、本来、個々の被験者が有している固有の表情特徴が欠落してしまうことも懸念される。したがって、如何に被験者によらない共通の手法で被験者固有の特徴を抽出し、かつ、それらの特徴を考慮した表情認識モデルを構築できるかが重要な課題である。

第二に、感情を 6 基本表情に分類することの妥当性に関する課題が挙げられる。通常、純粋な基本表情は滅多に表れず、多くの場合、表情は様々な感情が混在した状態で表れる。また、顔部位の動きの多様性により、それぞれの表情パターンは一意に定まらず、多様なパターンを形成する。すなわち、視覚的に識別可能な全ての表情を分類するためには、必ずしも 6 つのカテゴリが最適とは限らず、被験者に応じた最適なカテゴリ数が存在すると考えられる。したがって、表情の分類は、カテゴリ数が未知の分類問題と仮定できることから、被験者ごとに最適なカテゴリ数を自動的に決定する手法について検討が必要である。

理想的な表情認識システムは、分類できる表情のパターンが可能な限り豊富であるべきと考える。そのためには、ユーザの主観の入らない状態で表情パターンをカテゴリライズし、そのカテゴリに対してユーザが一意的に感情の意味づけを行うことが望ましい。すなわち、被験者によって感情のカテゴリは普遍的であっても表情のパターンは異なることから、システムはユーザの主観的な解釈により表情分類の粒度を変更可能であるべきと考える。

そこで、本論文では、表情の分類を未知のカテゴリ分類問題と仮定し、教師なし学習の自己組織化マップ (Self-Organizing Map, 以下, SOM と記す) [18] と教師あり学習のカウンタプロパゲーションネットワーク (Counter Propagation Network, 以下, CPN と記す) [19] を併用した被験者固有の表情特徴空間の生成手法を提案する。はじめに、表情の表出プロセスで発生する複雑な顔パターンの位相変化を、狭い写像空間の SOM を用いて階層的に学習し、被験者固有の表情カテゴリ数および各表情を特徴付ける顔パターンの代表画像 (基準ベクトル) を生成する。次に、各表情の代表画像を広い写像空間の CPN に入力することにより、表情の位相変化を捉えた顔パターンのカテゴリマップを生成し、無表情と基本 6 表情を基に各カテゴリに心理学的な意味付けを行う。なお、本論文では、このカテゴリ

マップを表情特徴空間として定義する。

本論文では、6 名の被験者を対象とした実験結果から、狭い写像空間の SOM を階層的に用いることにより、被験者固有の表情カテゴリ数および各カテゴリの代表画像を抽出可能であることを示す。また、CPN によって得られた表情特徴空間は、各表情カテゴリのクラスタの広がり表現しており、被験者固有の表情の豊かさを視覚的に分析可能であることを示す。

## 2. 提案手法

提案手法は、SOM を用いた学習データの正規化フェーズと CPN を用いた表情特徴空間の生成フェーズから構成される。

1. で述べたように、表情は多様なパターンを形成し、かつ、個人差を有することから、表情分類問題は未知のカテゴリ分類問題と考えることができる。したがって、大量の顔画像データの中から被験者固有の特徴に基づいた表情カテゴリを導出することが重要である。そこで、提案手法では、狭い写像空間を有する SOM の情報圧縮機能に着目した。SOM は教師なし学習アルゴリズムであることから、未知のカテゴリ分類問題に適している。学習データの正規化フェーズでは、SOM を階層的に用いることにより、表情の分類を 2 分木問題として扱い、得られた階層構造から表情画像群のカテゴリ数および各カテゴリの代表画像 (基準ベクトル) を生成した。次に、得られたカテゴリに対し、オペレータが意味付けを行うことにより、カテゴリを教師信号、代表画像を入力画像とする学習データを生成した。

正規化された学習データは、各カテゴリに含まれる表情画像群の代表画像であるため、典型的な表情パターンを表している。一方、表情認識を行う場合には、表情の表出の程度および中間表情への対応などといった汎化能力が大きな課題となる。したがって、各表情カテゴリのクラスタの広がりおよび隣接関係を表現した特徴空間を生成することが重要である。そこで、提案手法では、広い写像空間を有する CPN の情報拡張機能に着目した。表情特徴空間の生成フェーズでは、前述の処理により正規化された学習データ (教師信号、入力画像) を CPN の学習データとし、広い写像空間上にカテゴリマップを生成した。マップ上には、CPN の情報拡張機能により、学習データの連続性を補完するベクトルが生成される。カテゴリマップ上の各カテゴリの広さや隣接関係は、各被験者の特徴に基づいており、それぞれ異なると考えられることから、本論文では、このカテゴリマップを被験者固有の表情特徴空間として定義した。

以下、SOM、CPN のアルゴリズムについて述べた後、提案手法の処理手順について詳述する。

### 2.1 自己組織化マップ (SOM)

Kohonen によって提案された SOM は、人の脳が持つ自己組織化と適応学習をモデルとした学習アルゴリズムである [18]。SOM の構造は、学習データを入力する入力層と競合学習により自己写像を行う Kohonen 層の 2 層から構成される。SOM は、入力空間上に提示された特徴量から自己学習 (教師なし学習) を行い、入力空間の位相情報を内部の結合荷重として保持し、Kohonen 層に写像する。Kohonen 層のユニット間に結合関係は存在しないが、近傍学習の概念により、隣接するユニッ

トの結合荷重は学習過程で互いに類似した関係へ導かれる。

SOM の学習アルゴリズムを以下に示す。

(1)  $w_{i,j}(t)$  を時刻  $t$  における入力層ユニット  $i$  から Kohonen 層ユニット  $j$  への結合荷重とする。なお、 $w_{i,j}$  は乱数を用いて初期化している。

(2)  $x_i(t)$  を時刻  $t$  における入力層ユニット  $i$  への入力データとし、 $x_i(t)$  と  $w_{i,j}(t)$  のユークリッド距離  $d_j$  を (1) 式を用いて計算する。

$$d_j = \sqrt{\sum_{i=1}^I (x_i(t) - w_{i,j}(t))^2} \quad (1)$$

(3)  $d_j$  が最小となる Kohonen 層ユニットを探索する (これを勝者ユニットと称する)。

(4) 勝者ユニットの近傍領域  $N_c(t)$  に含まれる Kohonen 層ユニットの結合荷重  $w_{i,j}(t)$  を (2) 式を用いて更新する。 $\alpha(t)$  は学習率係数である。

$$w_{i,j}(t+1) = w_{i,j}(t) + \alpha(t)(x_i(t) - w_{i,j}(t)) \quad (2)$$

(5) (2) から (4) の処理を最大学習回数まで繰り返す。

提案手法では、Kohonen 層が 5 ユニットの 1 次元 SOM を採用した。学習回数  $t$  は、経験的に 200,000 回とし、近傍領域  $N_c(t)$  は勝者ユニットの第一近傍で固定とした。 $\alpha(t)$  の初期値は 0.5 とし、学習回数が 100,000 回の時点で 0.02、最大学習回数の時点で 0 までそれぞれ線形減少する構成とした。また、結合荷重の更新比率は勝者ユニットを 1 とし、近傍ユニットを 0.5 とした。

## 2.2 カウンタプロパゲーションネットワーク (CPN)

Nielsen によって提案された CPN は、2.1 で述べた Kohonen 学習則 (SOM) に Grossberg 学習則を組み合わせた学習アルゴリズムである。CPN の構造は、学習データを入力する入力層と、競合学習により自己写像を行う Kohonen 層、教師信号を対向伝搬して Kohonen 層に意味付けを行う Grossberg 層の 3 層から構成される。SOM は、学習データの位相特性に基づいた分類が可能であるが、教師なし学習アルゴリズムであるため、学習後の Kohonen 層の意味付け (ラベル付け) はオペレータによって行われる。一方、CPN は、学習データがどのカテゴリに属するかが決まっている場合に、Kohonen 層のラベルを自動的に決定する手法として有用である。このラベル付けされた Kohonen 層をカテゴリマップと呼ぶ。提案手法では、SOM を用いて正規化された学習データを基に、CPN による学習を行い、得られたカテゴリマップを被験者固有の表情特徴空間とした。

CPN の学習アルゴリズムを以下に示す。

(1)  $w_{n,m}^i(t)$  を時刻  $t$  における入力層ユニット  $i$  から Kohonen 層ユニット  $(n, m)$  への結合荷重、 $w_{n,m}^j(t)$  を時刻  $t$  における Grossberg 層ユニット  $j$  から Kohonen 層ユニット  $(n, m)$  への結合荷重とする。なお、 $w_{n,m}^i, w_{n,m}^j$  は乱数を用いて初期化している。

(2)  $x_i(t)$  を時刻  $t$  における入力層ユニット  $i$  への入力データとし、 $x_i(t)$  と  $w_{n,m}^i(t)$  のユークリッド距離  $d_{n,m}$  を (3) 式を

用いて計算する。

$$d_{n,m} = \sqrt{\sum_{i=1}^I (x_i(t) - w_{n,m}^i(t))^2} \quad (3)$$

(3)  $d_{n,m}$  が最小となる Kohonen 層ユニットを探索する (これを勝者ユニットと称する)。

(4) 勝者ユニットの近傍領域  $N_c(t)$  に含まれる Kohonen 層ユニットの結合荷重  $w_{n,m}^i(t)$ 、 $w_{n,m}^j(t)$  を (4) 式、(5) 式を用いて更新する。 $\alpha(t)$ 、 $\beta(t)$  は学習率係数、 $t_j(t)$  は時刻  $t$  における Grossberg 層ユニット  $j$  への教師信号である。

$$w_{n,m}^i(t+1) = w_{n,m}^i(t) + \alpha(t)(x_i(t) - w_{n,m}^i(t)) \quad (4)$$

$$w_{n,m}^j(t+1) = w_{n,m}^j(t) + \beta(t)(t_j(t) - w_{n,m}^j(t)) \quad (5)$$

(5) (2) から (4) の処理を最大学習回数まで繰り返す。

(6) 学習終了後、Kohonen 層の各ユニットから見た結合荷重  $w_{n,m}^j$  を比較し、最大値を示す Grossberg 層の教師 (カテゴリ) をそのユニットのラベルとする。

提案手法では、Kohonen 層を  $30 \times 30$  ユニットの Grossberg 層を学習データの正規化フェーズ (SOM) で得られたカテゴリ数とする CPN を採用した。学習回数  $t$  は経験的に 20,000 回とした。近傍領域  $N_c(t)$  の半径は初期値を Kohonen 層の 1/2 倍のサイズとし、学習回数が 10,000 回の時点まで線形的に減少させ、以降の 10,000 回は勝者ユニットの第一近傍で固定とした。 $\alpha(t)$ 、 $\beta(t)$  の初期値は 0.5 とし、学習回数が 10,000 回の時点で 0.02、最大学習回数の時点で 0 までそれぞれ線形減少する構成とした。また、結合荷重の更新比率は勝者ユニットを 1 とし、近傍ユニットを 0.5 とした。

## 2.3 表情特徴空間の生成手法

### 2.3.1 SOM の階層構造による表情特徴の正規化

SOM を用いた学習データの正規化フェーズに関する処理手順を図 1 に示す。また、処理の詳細を以下に示す。

(1) 取得した  $N$  枚の表情画像を SOM の学習データとした。

(2) Kohonen 層が 5 ユニットの 1 次元 SOM を用いて (1) の学習データの表情位相特性を学習した。これは、狭い写像空間に入力群の位相情報を圧縮することにより、SOM は与えられた入力群を大別する特徴的な表情位相特性を学習すると考えたためである。表情の位相構造は画像の輝度分布として表現されることから、入力層への入力値は画像の輝度分布とした。入力層のユニット数は、入力画像のサイズと同一である (図 1(a))。

(3) Kohonen 層の 5 ユニットの結合荷重  $w_{i,j}$  ( $0 \leq w_{i,j} \leq 1$ ) を 0 から 255 の値に変換し、結合荷重の可視化画像を作成した (図 1(b))。その後、5 枚の可視化画像列を目視により確認し、いずれの画像も同一の表情を表していると判断された場合、ひとつのカテゴリとみなして以降の階層化処理を終了した。また、この 5 枚の可視化画像列をこのカテゴリにおける代表画像として定義した。可視化画像列に複数の表情が混在していると認められた場合は、以降の階層化処理を継続した。

(4) Kohonen 層の各隣接ユニット間において結合荷重  $w_{i,j}$

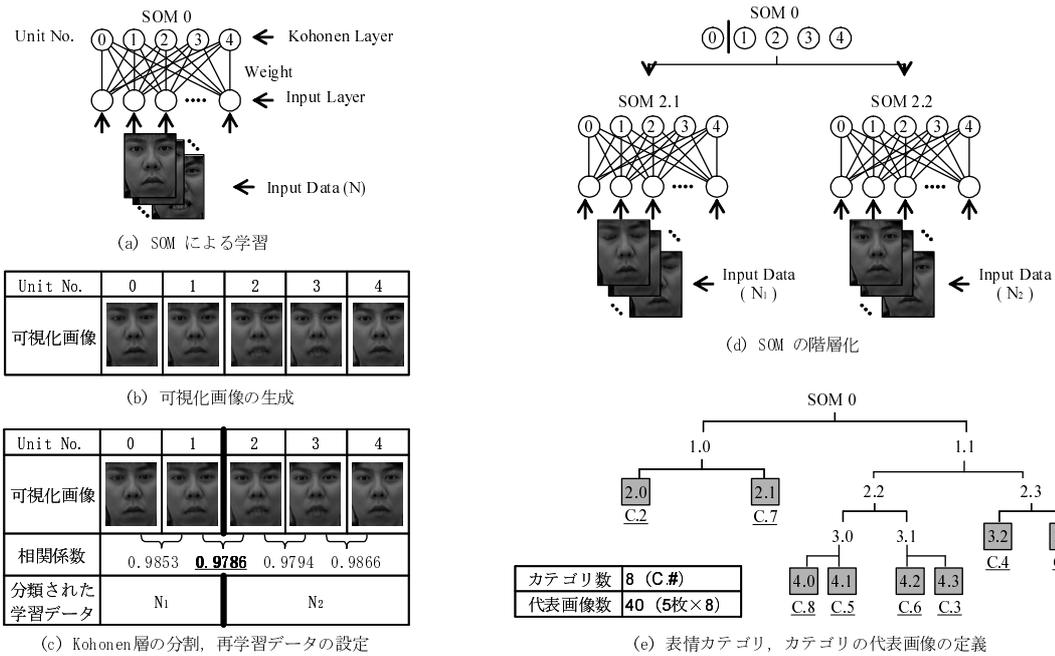


図 1 学習データ正規化フェーズの処理手順  
Fig. 1 Normalization process of training data.

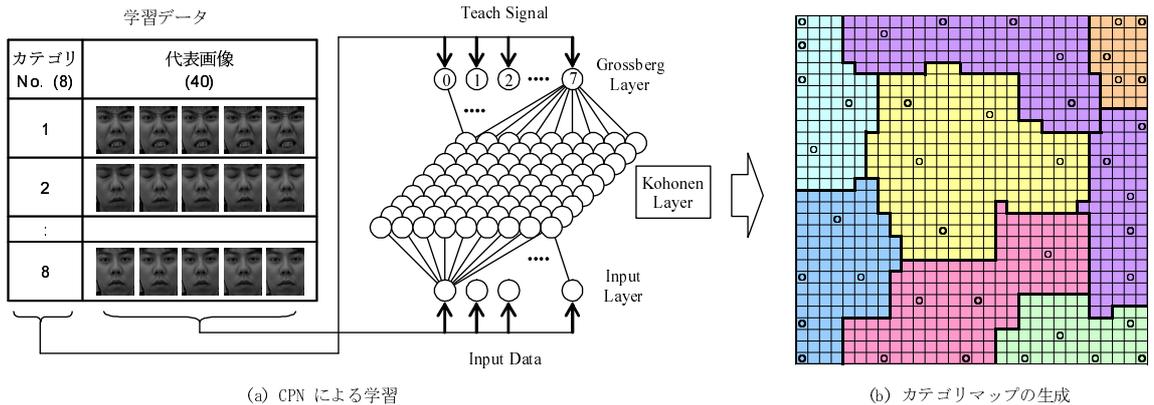


図 2 表情特徴空間生成フェーズの処理手順  
Fig. 2 Forming process of the expression feature space.

の相関係数を算出し、最小値のユニット間を境界として Kohonen 層を 2 分割した (図 1(c))。これは、SOM の近傍学習により、隣接するユニットの結合荷重は近い値へ導かれることから、境界の双方に分類された入力群は、大きな位相特性の差を有していると考えたためである。

(5) 2 分割した Kohonen 層の双方に分類された入力群 ( $N_1$  および  $N_2$ , ただし,  $N = N_1 + N_2$ ) は、それぞれ独立な部分問題と考えることができる (図 1(c))。そこで、双方に分類された入力群を対象として再帰的に (2) から (4) の処理を繰り返す、SOM の階層化を行った (図 1(d))。

(6) 上述の処理を繰返すことにより、表情分類の 2 分木構造を生成し、被験者固有の表情カテゴリ数および各カテゴリの代表画像を抽出した (図 1(e))。

### 2.3.2 CPN による表情特徴空間の生成

CPN を用いた表情特徴空間の生成フェーズに関する処理手

順を図 2 に示す。また、処理の詳細を以下に示す。

(1) 2.3.1 の 2 分木構造から得られたカテゴリおよび代表画像を、教師信号および入力データとし、CPN の学習データとした (図 2(a))。

(2) Kohonen 層を  $30 \times 30$  ユニット、Grossberg 層のユニット数を 2.3.1 で得られたカテゴリ数とする CPN を用いて、入力群の表情位相特性を学習した (図 2(a))。入力層への入力値は 2.3.1 と同様に代表画像の輝度分布とした。Grossberg 層への教師信号は、カテゴリを意味するユニットに 1 を、それ以外は 0 を入力した。入力層のユニット数は、入力画像のサイズと同一である。上述の処理を最大学習回数に達するまで繰返した。

(3) 学習終了後、Kohonen 層の各ユニットごとに Grossberg 層の各ユニットと連結された結合荷重を比較し、値の最も高いカテゴリをそのユニットのラベルとして定義した。

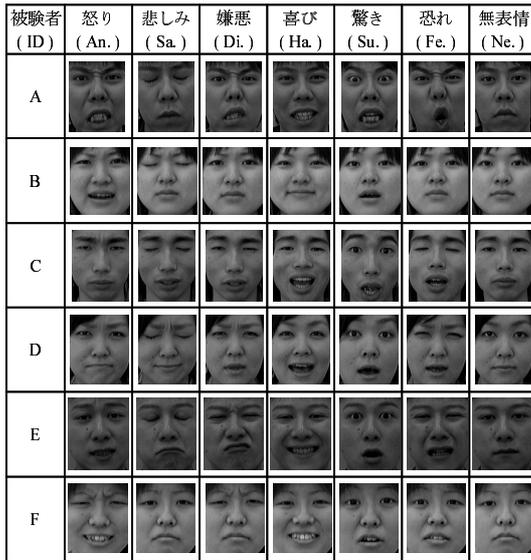


図3 表情画像

Fig. 3 Facial expression images.

(4) 上述の処理によって得られたカテゴリマップを被験者固有の表情特徴空間として定義した(図2(b)). なお, カテゴリに対する心理学的な意味づけは, 無表情と基本6表情に基づいて行い, 基本6表情に属しないと判断されたカテゴリは全て無表情のカテゴリとして定義した.

### 3. 対象とする表情画像

本論文で使用した表情画像の一例を図3に示す. 対象とする表情は, 6名の被験者が主観的に表現した6基本表情と無表情の7表情とした. 撮影環境は, 通常室内(日常一般的と考えられる蛍光灯による照明下)とし, 被験者の頭部がフレーム中に含まれる状態で正面顔画像を撮影した.

無表情と表情を交互に5回ずつ表出し(無-表情-無-表情-無-表情-無-表情-無-表情-無), 10 frame/secの動画像で取得した後, グレースケール(8bit)の静止画像に変換した. また, 無表情のみを10秒間(100フレーム)取得した. 全被験者の学習データのフレーム数を表1に示す. 代表的な表情分析モデルであるFACSでは, 眉や目周辺, 鼻や口周辺といった顔部位の動き特徴がAUとして定性的に定義されている. したがって, 対象とする領域はこれらの顔部位を含む顔全体領域とし, 各フレームから切り出して使用した.

なお, 多くの文献では, 一般に公開されている顔表情データベース(例えば[7], [8])を用いた評価実験が行われている. これらのデータベースでは, 一被験者, 一表情につき数枚の画像が提供されている. 提案手法では, 大量の表情画像群から被験者固有の表情カテゴリとその代表画像を抽出する. したがって, 提案手法の評価としては適さないと考えられるため, 本論文では独自に取得した表情画像を使用した.

### 4. 提案手法の適用結果および考察

#### 4.1 SOMによる表情特徴の正規化に関する考察

被験者6名に対し, 提案手法を適用して生成したSOMの2

表1 学習データのフレーム数

Table 1 The number of training data.

ID	An.	Sa.	Di.	Ha.	Su.	Fe.	Ne.	Total
A	136	198	143	169	127	140	100	1013
B	152	136	153	162	154	190	100	1047
C	192	173	154	158	153	156	100	1086
D	152	158	178	177	158	170	100	1093
E	95	113	108	112	109	108	100	745
F	165	197	198	163	165	167	100	1155

分木構造を図4に示す. また, 図4から抽出した表情のカテゴリ数および代表画像数を表2に示す. 図4より, 2分木構造は被験者ごとに異なっている様子が認められる. また, 表2より, 全被験者に対して基本6表情と無表情の7カテゴリを学習対象としたにもかかわらず, SOMの自己写像特性により, 被験者ごとにそれぞれ異なるカテゴリ数が導出されている様子が認められる.

被験者Aでは, 8カテゴリが生成されており, 代表画像が40枚抽出されている. このことは, 被験者Aでは, 学習データ中の表情パターンが安定しており, 6基本表情がそれぞれ1つのカテゴリ, 無表情が2つのカテゴリとして生成されたことを示している.

また, 被験者Dでは, 15のカテゴリが生成されており, 代表画像が75枚抽出されている. ここで, 喜びに着目すると, 同一表情にも係らず, 3つのカテゴリが生成されている. 図5に被験者Dの喜びの代表画像を示す. 図5より, 目を閉じて歯を見せる喜び(同図(a)), 微笑み(同図(b)), 口を大きく開く喜び(同図(c))の3パターンが生成されている様子が認められる. このことは, 被験者Dの喜びの表情が, 複数の表情パターンを有していることを示しており, SOMの2分木構造では, それぞれを異なる表情位相特性として学習し, 別カテゴリに分類したことを示している.

一方, 被験者Bでは, 7カテゴリが生成されており, 代表画像が35枚抽出されている. ここで, 嫌悪と恐れに着目すると, 両者が1つのカテゴリとして生成されている. 図3に示した被験者Bの嫌悪と恐れを比較すると, 表情のパターンが類似しており, 視覚的に判別が困難であることがわかる. このことは, SOMの2分木構造が表情の位相特性の類似した表情をひとつのカテゴリとして生成したことを示している.

以上より, 提案手法を用いることにより, 同一表情であっても複数のカテゴリが存在する場合は, それぞれ別のカテゴリとして定義できること, また, 視覚的に類似した表情は1つのカテゴリとして定義できることが明らかとなった. なお, 本論文では, 2分木構造より, 全てのカテゴリに対して意味づけを行ったが, ユーザの主観に合わせ, 注目したいカテゴリのみを意味づけする, または, 同一表情であっても, 階層化を進めることにより, カテゴリの細分化を行うことも可能である.

例えば, 被験者Eでは, 図6(a)に示すように第4階層の可視化画像が全て驚きの表情を示している. 本論文では, この段階で, 被験者Dの驚きのカテゴリを決定し, 可視化画像を代表

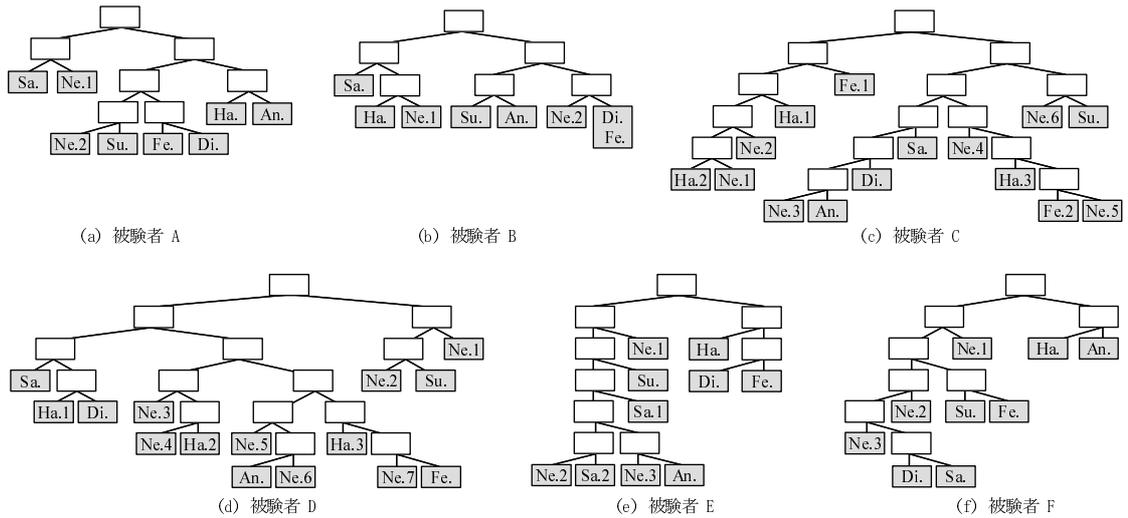


図 4 提案手法により生成された SOM の 2 分木構造

Fig. 4 Results of hierarchical SOM structure.

表 2 表情カテゴリ数および代表画像数

Table 2 The number of expression categories and representative images.

ID	カテゴリ	An.	Sa.	Di.	Ha.	Su.	Fe.	Ne.	画像
A	8	1	1	1	1	1	1	2	40
B	7	1	1	1	1	1	*	2	35
C	15	1	1	1	3	1	2	6	75
D	15	1	1	1	3	1	1	7	75
E	10	1	2	1	1	1	1	3	50
F	9	1	1	1	1	1	1	3	45

\* Di. と同一カテゴリ

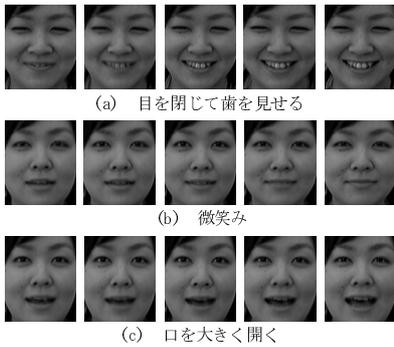


図 5 カテゴリ代表画像 (被験者 D の喜び)

Fig. 5 Representative category images (Happiness of subject D).

画像として抽出した。しかしながら、階層化を一階層進めることにより、「眉を大きく上げる驚き (図 6(b))」、「眉を小さく上げる驚き (図 6(c))」、さらに一階層進めることにより、「口を小さく開く驚き (図 6(d) および (f))」、「口を大きく開く驚き (図 6(e) および (g))」というように、同一表情であっても表情パターンの微小な変化による詳細化したカテゴリを生成することが可能である。

#### 4.2 CPN による表情特徴空間の生成に関する考察

4.1 で抽出したカテゴリおよび代表画像 (表 2) を、CPN の

教師信号および入力として生成した表情特徴空間を図 7 に示す。図中の丸印のユニットは、学習後の CPN に学習データを入力した際に発火したユニットを示している。同図より、表情特徴空間上の表情カテゴリの広さ (ラベル数) が被験者ごとに異なっている様子が認められる。また、一被験者に注目した場合でも、それぞれの表情カテゴリのラベル数に差が認められる。

被験者ごとのラベル数の割合を表 3 に示す。被験者 A では、悲しみが 4.1%、嫌悪が 25.2% であり、同じ 5 枚の学習データを使用したにもかかわらず、ラベル数に大きな差が認められる。また、図 7(a) より、悲しみは学習データの発火ユニットが密集しており、嫌悪は広く分散している様子が認められる。このことは、悲しみでは、使用した 5 枚の学習データの位相特性が他の表情と比較して類似しており、表情パターンの変化が小さいこと、また、嫌悪では、学習データの位相特性の差が大きく、表情パターンの変化が大きいことを示していると考えられる。同様に被験者 D では、3 カテゴリで生成された喜びの表情が豊かであり (15.6%)、驚きの表情は変化の少ないことが認められる (3.0%)。また、被験者 F では、恐れ表情が豊かであり (20.2%)、嫌悪の表情は変化の少ないことが認められる (3.3%)。

さらに、被験者 E の表情特徴空間の驚きの部分を拡大し、各ユニットの結合荷重を可視化した結果を図 8 に示す。同図より、5 枚の学習データの発火ユニット周辺では、微小にパターンの異なる表情が生成されている様子が認められる。提案手法における CPN は、カテゴリごとに同数の学習データ (1 カテゴリにつき 5 枚) を使用し、広い 2 次元空間 ( $30 \times 30 = 900$  ユニット) へ写像を行うことから、情報の拡張機能を利用した処理と言える。すなわち、写像空間では、近傍学習により 5 つの代表画像の連続性を補完するユニットが生成される。したがって、表情特徴空間上のラベル数は、カテゴリ内での位相特性の差の程度、すなわち、被験者固有の表情の豊かさを表現しているものと考えられる。

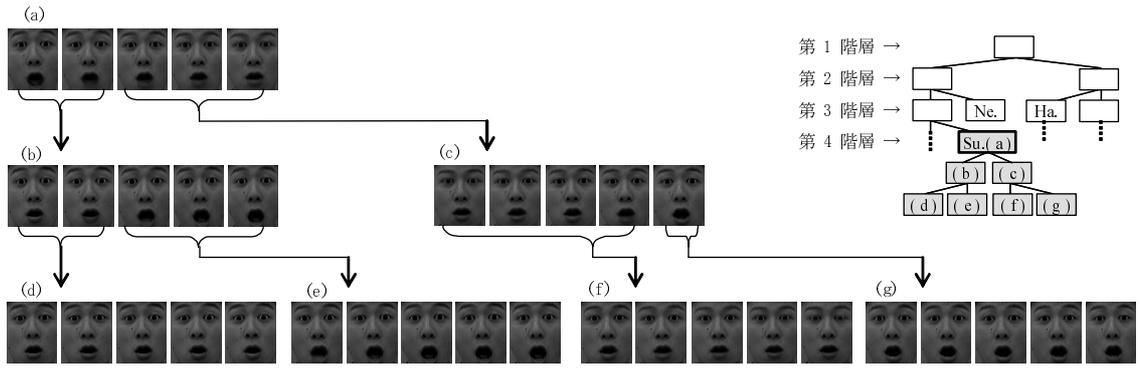


図 6 表情カテゴリの細分化 (被験者 E の驚き)

Fig. 6 Subdivision of an expression category (Surprise of subject E).

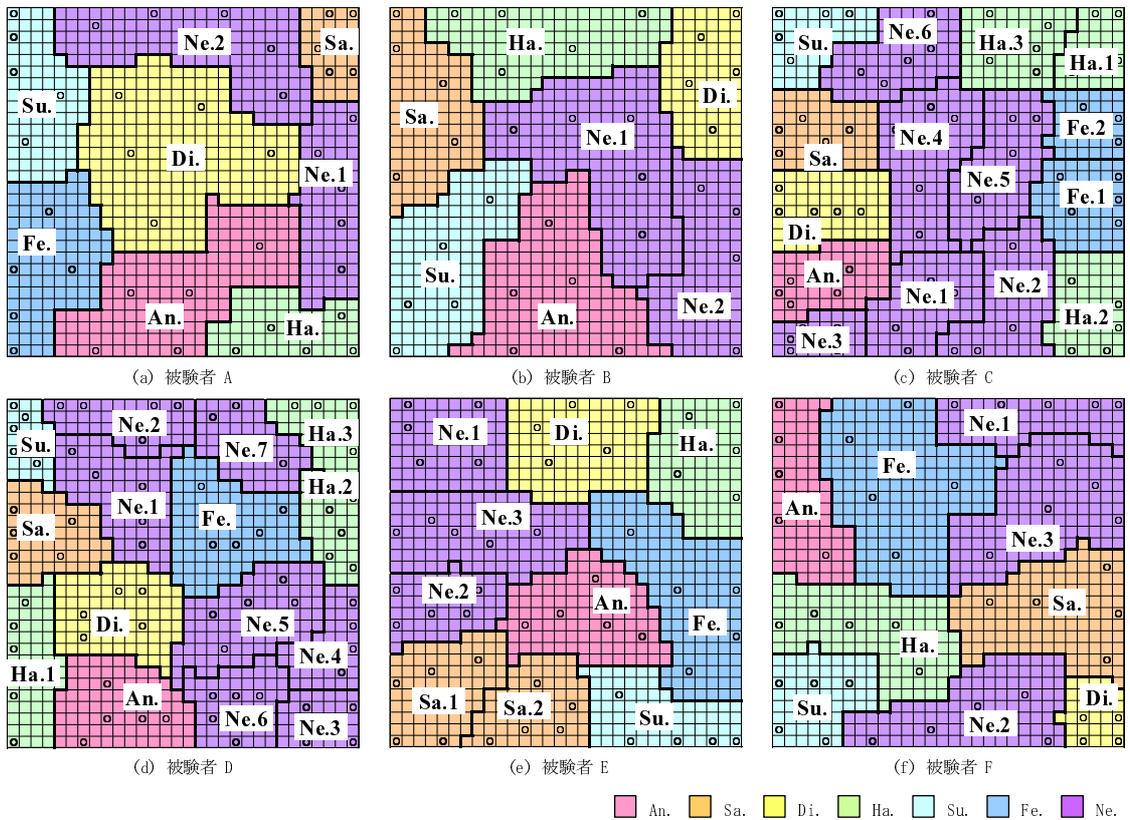


図 7 提案手法により生成された表情特徴空間

Fig. 7 Results of the expression feature space.

## 5. まとめ

本論文では、表情を未知のカテゴリ分類問題と仮定し、被験者固有の表情特徴空間の生成手法について検討を加えた。本論文で得られた知見を以下に示す。

- 狭い写像空間の SOM を階層的に用いる (2 分木構造) ことにより、被験者固有の表情のカテゴリ数およびカテゴリの代表画像を抽出できることを示した。また、2 分木構造の可視化画像より、各カテゴリに対し、典型的な表情、混在した表情などといった被験者独自の意味づけを行い得ること、加えて、2 分木構造の階層数を増加することにより、顔部位の微小な形状変化によるカテゴリの細分化を行い得ることを示した。提案

手法では、オペレータの主観の入らない状態で自己組織的に表情画像の位相特性を学習し、被験者固有の表情カテゴリおよびカテゴリの代表画像を生成する。またオペレータの解釈により、カテゴリの粒度、意味づけを設定できることから、表情を未知の分類問題と仮定した場合の学習データの正規化手法として有用であることが示唆された。

- 2 分木構造より得られたカテゴリ数およびカテゴリ代表画像を、広い写像空間を有する CPN の学習データとして使用したところ、被験者固有の表情の豊かさが表情特徴空間上のラベル数で表現できることが明らかとなった。このことは、CPN の情報拡張機能により、表情特徴空間上で被験者固有の表情カテゴリの定量化が行われていることを示唆しており、提案手法に

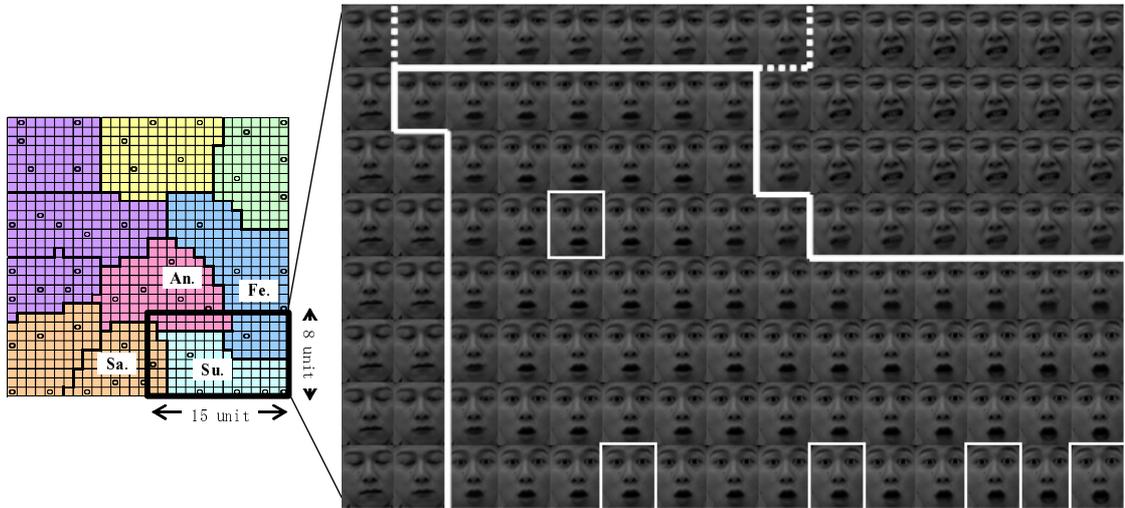


図 8 表情特徴空間の拡大図 (被験者 E の驚き)

Fig. 8 Magnified view of the expression feature space (Surprise of subject E).

表 3 表情特徴空間上のラベル数の割合

Table 3 Ratio of the number of labels on the expression feature space.

単位 : %

ID	An.	Sa.	Di.	Ha.	Su.	Fe.	Ne.
A	17.7	4.1	25.2	7.2	9.4	11.9	24.4
B	18.0	11.7	9.2	13.3	14.8	*	33.0
C	6.3	6.6	7.1	17.1	5.0	10.7	47.2
D	9.1	6.8	9.2	15.6	3.0	11.0	45.3
E	11.6	16.2	11.6	9.4	8.0	15.7	27.6
F	9.1	13.9	3.3	12.8	7.1	20.2	33.6

\* Di. と同一カテゴリ

よって生成された特徴空間を表情認識の識別器として使用した場合、汎化能力の向上に寄与することが期待される。

なお、本論文では、表情特徴空間の生成手法の検討に焦点を当てたため、固定した照明環境下および同時期に取得した表情画像を使用した。今後は、照明条件に対する頑強性の向上を目的とし、前処理として Gabor Wavelet などを施した表情画像を用いることにより、提案手法の検証を行う。また、長期間継続して取得した表情画像を用いることにより、表情の経時変化に関する検証を行う予定である。

#### 文 献

[1] B. Fasel and J. Luettin, "Automatic Facial Expression Analysis: A Survey," *Pattern Recognition*, Vol.36, pp.259-275, 2003.

[2] M. Yang, D. J. Kriegman, and N. Ahuja, "Detecting Faces in Images: A Survey," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.24, No.1, pp.34-58, 2002.

[3] M. Pantic and L. J.M. Rothkrantz, "Automatic Analysis of Facial Expressions: The State of the Art," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.22, No.12, pp.1424-1445, 2000.

[4] W. Zhao, R. Chellappa, A. Rosenfeld, and P.J. Phillips, "Face Recognition: A Literature Survey," Technical Report 00-948, University of Maryland, 2000.

[5] 長谷川修, 森島繁生, 金子正秀, "「顔」の情報処理," *信学論*

(D-II), Vol.J80-D-II, No.8, pp.2047-2065, 1997.

[6] M. Pantic and L. J.M. Rothkrantz, "Expert System for Automatic Analysis of Facial Expression," *Image and Vision Computing*, Vol.18, No.11, pp.881-905, 2000.

[7] M.J. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding Facial Expressions with GaborWavelets," *Proc. Third IEEE International Conference on Automatic Face and Gesture Recognition*, pp.200-205, 1998.

[8] M.J. Lyons, J. Budynek, and S. Akamatsu, "Automatic Classification of Single Facial Images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.21, No.12, pp.1357-1362, 1999.

[9] Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu, "Comparison Between Geometry-Based and Gabor-Wavelet-Based Facial Expression Recognition Using Multi-Layer Perceptron," *Proc. Third IEEE International Conference on Automatic Face and Gesture Recognition*, pp.454-459, 1998.

[10] 篠原雄介, 大津展之, "フィッシャー重みマップを用いた顔画像からの表情認識," *信学技報*, PRMU2003-269, 2004.

[11] 伊達章, 倉田耕治, "連続自己組織化マップによる表情と個人情報の分離," *画像の認識・理解シンポジウム (MIRU2004) 論文集 I*, pp.613-617, 2004.

[12] 肖業貴, N.P. チャンドラシリ, 田所嘉昭, 尾田政臣, "2-D DCT とニューラルネットワークを用いた顔画像の表情認識," *信学論 (A)*, vol.J81-A, no.7, pp.1077-1086, Jul.1998.

[13] 下田宏, 國弘威, 吉川榮和, "動的顔画像からのリアルタイム表情認識システムの試作," *ヒューマンインタフェース学会論文誌*, Vol.1, No.2, pp.25-32, 1999.

[14] 大塚尚弘, 大谷淳, 中津良平, "連続出力確率密度分布を用いた HMM による動画からの複数人物の表情認識," *信学論 (D-II)*, Vol.J80-D-II, No.8, pp.2129-2137, 1997.

[15] 海老根秀之, 池田美, 中村納, "基準表情からの顔部位変化の追跡に基づく表情認識," *電学論 C*, Vol.122-C, No.3, pp.471-484, 2002.

[16] P. Ekman and W.V. Friesen, "Facial Action Coding System," Consulting Psychologist Press, 1978.

[17] P. Ekman and W.V. Friesen, "Unmasking the face," Prentice Hall, Inc., 1975. (工藤訳, "表情分析入門," 誠信書房, 1987.)

[18] T. Kohonen, "Self-organizing maps," Springer Series in Information Sciences, 1995.

[19] R. H. Nielsen, "Counterpropagation networks," *Applied Optics*, vol.26, No.23, pp.4979-4984, 1987.