

推薦論文

# 適応学習機能を有する表情特徴空間の生成と 時間軸に対する頑強性の評価

石井 雅樹<sup>1,a)</sup> 山屋 孝史<sup>2,†1</sup> 景山 陽一<sup>2</sup> 高橋 毅<sup>2</sup> 西田 眞<sup>2,†2</sup>

受付日 2016年12月3日, 採録日 2017年4月10日

**概要:** 近年, 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注目を集めている. 従来提案されている表情認識モデルは, 短期間に撮影された表情画像を学習データとして識別器を生成している場合が多い. しかし, 人間が表出する表情パターンは多様であるため, 短期間ですべての表情パターンを学習データとして取得することは困難である. 時々刻々と変化する表情には, 学習データに含まれていない表情パターンが多数出現する. このため, 初期に生成した識別器を用いて, これらの表情を長期的に安定して認識することは困難と考えられる. 著者らは, 表情認識モデルが長期的な時間軸に対して高い頑強性を維持していくためには, 初期に生成した識別器が時間経過とともに進化・適応していくべきと考えている. すなわち, 既存の知識 (過去に学習した表情パターン) を保持したまま, 新たな知識 (未学習の表情パターン) を追加学習することが可能な表情認識モデルの確立が必要と考える. 本論文では, 上記に示した適応的な学習機能を有する表情認識モデルの生成手法を提案する. また, 表情認識モデルの追加学習機能に焦点を絞った評価実験の結果から, 提案手法の有用性について述べる.

**キーワード:** 表情認識, 適応共鳴理論, 適応学習, 対向伝播ネットワーク

## Generation of Expression Feature Space with Adaptive Learning Capability and Evaluation of Robustness along a Time Axis

MASAKI ISHII<sup>1,a)</sup> TAKASHI YAMAYA<sup>2,†1</sup> YOICHI KAGEYAMA<sup>2</sup> TSUYOSHI TAKAHASHI<sup>2</sup>  
MAKOTO NISHIDA<sup>2,†2</sup>

Received: December 3, 2016, Accepted: April 10, 2017

**Abstract:** Most facial expression recognition models eventually create a classifier based on facial expression images taken during a short period of time, and use them as the base data for learning. However, because so many facial expression patterns exist, representations cannot be made of all of them, and it is difficult to obtain all available patterns within a short period of time and retain them for use as learning data. Actual facial expressions that change from one moment to the next will at times create different patterns not contained in the learning data. For a facial expression recognition model to retain its high robustness continuously along the time axis over a long period of time, the classifier created during the initial stage should evolve and gradually become adaptive over time. In other words, it is necessary for the model to retain its existing knowledge (i.e., previous facial patterns) and simultaneously learn by adding newly available knowledge (i.e., new facial patterns) as it becomes available. We propose a method for creating a facial expression recognition model that can offer such an adaptive learning capability. In addition, the usefulness of the proposed model is described based on the evaluation results of its incremental learning capability.

**Keywords:** facial expression recognition, adaptive resonance theory, adaptive learning capability, counter propagation networks

<sup>1</sup> 秋田県立大学システム科学技術学部  
Akita Prefectural University, Yurihonjyo, Akita 015-0055, Japan

<sup>2</sup> 秋田大学大学院理工学研究科  
Akita University, Gakuen-machi, Akita 010-8502, Japan

<sup>†1</sup> 現在, 日本ビジネスシステムズ株式会社  
Presently with Japan Business Systems, Inc.

<sup>†2</sup> 現在, 放送大学  
Presently with The Open University of Japan

a) ishii@akita-pu.ac.jp

## 1. はじめに

コンピュータが人間の行動から意図・感情を理解することができれば、状況に応じた判断・行動が可能になるため、知的なヒューマンマシンインターフェイスの実現が可能となる。我々人間のコミュニケーションは単語の意味の連続情報、文面などの言語情報のほかに、音声、表情、身振りなどの非言語情報を用いて行われる。中でも人間の顔表情は、顔の構成器官である眉、目、鼻、口などの形状、および、これらの動きにより表出され、多くの情報を有していることから、意思疎通において重要な役割を担っている。Ray 氏らの報告によれば、コミュニケーションにおいて言語情報は全体の 35% の割合で寄与しており、残りの 65% は非言語情報が占めるとされている [1]。また、Ekman 氏らは、表情に表出する基本的感情は異文化間でも共通性が高いものであると述べている [2]。したがって、ヒューマンマシン間の感情のコミュニケーションを実現するためには、顔表情に着目することが有用であると考えられる。

上記を目的とした顔表情の認識に関する研究は現在まで活発に行われており、これまでに提案された主な認識手法は、赤松氏や Pantic 氏らにより文献にまとめられている [3], [4], [5], [6], [7], [8]。また、近年では実際のコミュニケーションに沿った、より自然かつ微妙な表情の認識に関する研究報告もなされている [9], [10]。しかし、従来研究では、短期間で撮影した被験者の表情画像を用いて表情認識モデルを生成し、評価実験を行っている場合が多い。人間の表情は静的多様性（顔の構成器官の形状および顔全体における配置）と動的多様性（情動の発露および意思伝達時における顔面筋肉の動き）を有しているため、短期間の撮影で個人のすべての表情パターンを取得することは困難であると考えられる。時々刻々と変化する表情には、未学習の表情パターンが多数含まれている。したがって、短期間で取得した表情画像を用いて表情認識モデルを生成した場合は、時間経過によって出現する未学習の表情パターンに対して認識率が低下することが懸念される。

表情認識モデルが長期的な時間軸に対して高い頑強性を維持していくためには、初期に生成した識別器が時間経過とともに進化・適応していくべきと考えられる。すなわち、既存の知識（過去に学習した表情パターン）を保持したまま、新たな知識（未学習の表情パターン）を追加学習することが可能な表情認識モデルの確立が必要と考える [11]。著者らは先行研究において、表情から感情の程度を推定するための個人固有の表情特徴空間の生成手法を提案した [12], [13]。しかし、この特徴空間は未学習の表情画像に対する汎化能力を有しているものの、追加的な学習機能を有していない。小川氏らは 9 日間にわたって取得した表情画像を用いて、日々の学習の積み重ねによる表情の学習・認識の可能性を示している [14]。具体的には、異なる日時

に取得した入力情報（表情）が辞書情報（識別器）に積み重なることにより、環境に対する変化を吸収すること、すなわち、汎化能力が向上することを明らかにしている。しかし、上述の提案では、過去に取得したすべての表情画像を学習に用いることから、学習データは時間経過とともに増加し、追加学習を行っているとはいえない。

そこで本論文では、時間軸に対して頑強性を有する表情認識モデルの実現を目的とし、適応学習機能を有する表情認識モデルの生成手法を提案する。具体的には、安定性と可塑性に優れた追加学習アルゴリズムである適応共鳴理論 (Adaptive Resonance Theory: ART) [15] を上述の表情特徴空間 [12], [13] に連結し、適応的な学習を試みた。また、表情認識モデルの適応学習機能に焦点を絞った評価実験の結果から、提案手法の有用性について考察する。

本論文は 7 章より構成されている。2 章では先行研究で提案された表情特徴空間の生成手法について述べる。3 章では適応学習機能を有する表情特徴空間の生成手法について述べる。4 章では本論文で用いた表情画像の詳細を述べ、5 章で提案手法を用いた適応学習機能の基礎実験の結果について考察する。6 章では 11 日間にわたって取得した表情画像を用いた適応学習の評価実験の結果について考察する。7 章では本論文で得られた知見および今後の検討課題についてまとめる。

## 2. 個人固有の表情特徴空間の生成

### 2.1 概要

先行研究において、表情から感情の程度を推定するための個人固有の表情特徴空間の生成手法が提案されている [12], [13]。本章では先行研究の概要について述べる。上述の手法では対向伝播ネットワーク (Counter Propagation Networks: CPN) [16] を用いて表情画像を学習することにより、2 種類の表情特徴空間（表情空間マップ、感情空間マップ）を生成している。

表情特徴空間 (CPN の Kohonen 層) に表情画像を入力した場合、ユークリッド距離の最も小さい Kohonen 層上のユニット（勝者ユニット）にラベル付けされたカテゴリーが認識結果となる。具体的には、表情空間マップは入力された表情画像に対して、Ekman 氏らによって提唱された 6 基本感情（怒り、悲しみ、嫌悪、喜び、驚き、恐れ）[2] と無表情の感情カテゴリーを一意に決定する。また、感情空間マップは物理的な表情パターンの変化の程度に対して、「快-不快」および「覚醒度」を指標とした心理的な感情の程度を定量化することが可能である。しかし、上述の特徴空間は未学習の表情画像に対する汎化能力を有しているものの、追加的な学習機能を有していないのが現状である。

本論文の内容は 2015 年 12 月の支部研究会にて報告され、支部長により情報処理学会論文誌ジャーナルへの掲載が推薦された論文である。

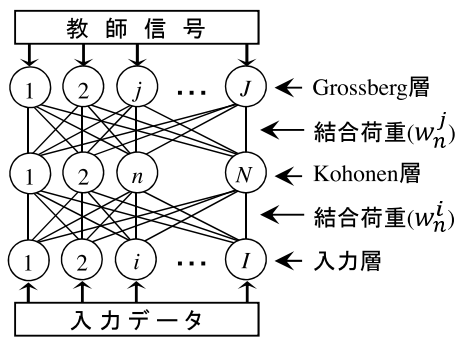


図 1 CPN の基本構造  
Fig. 1 CPN architecture.

## 2.2 対向伝播ネットワーク (CPN)

CPN は自己組織化マップ (Self-Organizing Maps: SOM) [17] に Grossberg 学習則を組み合わせた教師あり学習アルゴリズムであり、入力層、競合学習により自己写像を行う Kohonen 層、教師信号を対向伝播して Kohonen 層にラベル付けを行う Grossberg 層の 3 層から構成される。図 1 に CPN の基本構造を示す。また、CPN の処理手順を以下に示す。

- (1)  $w_n^i(t)$  を時刻  $t$  における入力層ユニット  $i$  から Kohonen 層ユニット  $n$  への結合荷重,  $w_n^j(t)$  を時刻  $t$  における Grossberg 層ユニット  $j$  から Kohonen 層ユニット  $n$  への結合荷重とする。なお,  $w_n^i(t)$ ,  $w_n^j(t)$  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  $w_n^i(t)$  のユークリッド距離  $d_n$  を式 (1) により算出する。

$$d_n = \sqrt{\sum_{i=1}^I (x_i(t) - w_n^i(t))^2} \quad (1)$$

- (3)  $d_n$  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  $w_n^i(t)$ ,  $w_n^j(t)$  を式 (2), 式 (3) を用いて更新する。  $\alpha(t)$ ,  $\beta(t)$  は学習率係数,  $x_j(t)$  は時刻  $t$  における Grossberg 層ユニット  $j$  への教師信号である。

$$w_n^i(t+1) = w_n^i(t) + \alpha(t)(x_i(t) - w_n^i(t)) \quad (2)$$

$$w_n^j(t+1) = w_n^j(t) + \beta(t)(x_j(t) - w_n^j(t)) \quad (3)$$

- (5) 前述の処理を最大学習回数まで繰り返す。
- (6) 学習終了後, Kohonen 層のユニットごとに Grossberg 層の結合荷重  $w_n^j(t)$  を比較し, 最大値を示す Grossberg 層ユニットのカテゴリ (教師信号) をそのユニットのラベルとする。

## 2.3 表情空間マップ生成手法

CPN を用いた表情空間マップの生成手法について説明

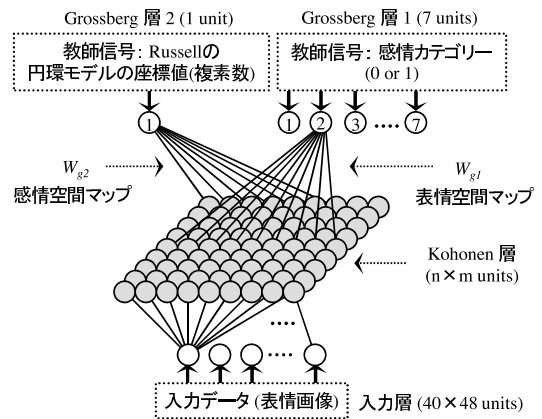


図 2 表情特徴空間の生成手法

Fig. 2 Generation method of facial expression feature space.

する。基本的な CPN との違いは、図 2 に示すように異なる教師信号を与える 2 種類の Grossberg 層を有することである。処理手順を以下に示す。

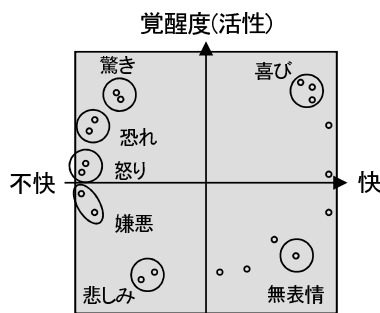
- (1) CPN の入力層のユニット数を  $40 \times 48$  ユニット (表情画像のサイズ), Kohonen 層のユニット数を任意, 学習回数は 20,000 回とした。教師信号を与える Grossberg 層は Ekman 氏らによって提唱された 6 基本感情 [2] と無表情を表すものとし, 7 ユニットとした (Grossberg 層 1)。
- (2) 表情画像を学習データとして入力層へ入力し, 学習を行った。Grossberg 層 1 への教師信号は, 入力データが属する感情カテゴリーユニットには 1 を, それ以外には 0 を入力した。また, 近傍領域  $N_c(t)$  の半径は, Kohonen 層のすべてのユニットが学習過程において必ず 1 回は更新されるように初期値を Kohonen 層の  $1/2$  倍のサイズとし, 最大学習回数の時点で第一近傍まで線形減少させた。学習係数  $\alpha(t)$ ,  $\beta(t)$  も同様に 0 まで線形減少させた。
- (3) 前述 (2) の処理を最大学習回数まで繰り返し行った。
- (4) 学習終了後, Kohonen 層のユニットごとに Grossberg 層 1 の結合荷重 ( $W_{g1}$ ) を比較し, 値の最も高い感情カテゴリーをそのユニットのラベルとした。前述の処理によって生成されたカテゴリーマップを個人固有の表情空間マップとして定義した。

生成した表情空間マップに表情画像を入力した場合, ユークリッド距離の最も小さいユニットにラベル付けされた感情カテゴリーが認識結果となる。

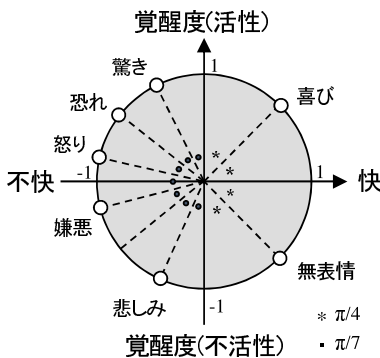
## 2.4 感情空間マップ生成手法

2.3 節の処理と並行し, Russell の円環モデル [18] に基づく座標値を CPN の教師信号 (Grossberg 層 2) として入力することにより, 表情パターンの変化の程度と感情の程度を対応付ける感情空間マップを生成した。Russell の円環モデルとは心理学の分野で提唱された心理空間モデルであ





(a) Russell の円環モデル



(b) CPN の教師信号

図 3 「快-不」および「覚醒度」を軸とした心理空間モデル

Fig. 3 Mental space model based on pleasure–displeasure and arousal axes.

る。なお、本論文で対象とする被験者の年齢は 20 歳代である。したがって、円環モデルは図 3 (a) に示す大人の心理空間モデルを採用した。以下に処理の手順を示す。

- (1) 図 3 (a) に示すように、Russell の円環モデルでは各表情刺激が「快-不快」および「覚醒度」を軸とした空間上で円環状に整列されている。提案手法ではこの円環状の空間を複素平面として表現し、Grossberg 層 2 の教師信号として同図 (b) に基づく複素数を入力した。たとえば、入力される学習データが喜びの感情カテゴリである場合、Grossberg 層 2 へ入力する教師信号は  $\cos(\pi/4) + i \sin(\pi/4)$  となる。
- (2) 前述の処理を 2.3 節の処理と並行して最大学習回数まで繰り返した。
- (3) 学習終了後、Grossberg 層 2 の結合荷重 ( $W_{g2}$ ) の実数部および虚数部の値をもとに、Kohonen 層の各ユニットを複素平面に布置した。また、この複素平面を個人固有の感情空間マップとして定義した。

感情空間マップに表情画像を入力した場合、表情パターンの物理的な変化の程度に応じて「快-不快」および「覚醒度」を指標とした感情の定量化が可能となる。

### 3. 適応学習機能を有する表情特徴空間の生成

#### 3.1 概要

2 章で述べた表情特徴空間は追加的な学習機能を有して

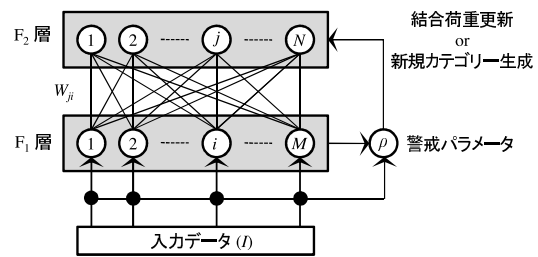


図 4 Fuzzy-ART の構造

Fig. 4 Fuzzy-ART architecture.

いないため、時間経過にともなう未学習データの出現により表情の認識性能が低下することが懸念される。そこで本論文では、安定性と可塑性に優れた追加学習アルゴリズムである適応共鳴理論 (Adaptive Resonance Theory: ART) [15] を特徴空間に連結することにより適応的な学習を試みた。

#### 3.2 適応共鳴理論 (ART)

ART は警戒パラメータと呼ばれる値を用いて入力データを既存の知識 (既存カテゴリ) と比較する。両者の整合度が警戒パラメータより高い場合、入力データは既存カテゴリに含まれると判断され、既存カテゴリの結合荷重が更新される。一方、整合度が警戒パラメータよりも低い場合、入力データは既存カテゴリに含まれない新たなカテゴリであると判断され、新規カテゴリとして入力データが登録される。ART には基本形の ART-1 をはじめとして様々な種類が存在するが、本論文では画像の輝度値を入力とするため、0~1 のアナログ入力に対応可能な Fuzzy-ART を採用した。Fuzzy-ART の基本構造を図 4 に示す。また、学習アルゴリズムの詳細を以下に示す。

- (1)  $I(t) \in [0, 1]_i$  を時刻  $t$  における入力層への入力データとする。
  - (2) 入力データ  $I(t)$  と各  $F_2$  層の各ユニット  $j$  の選択強度  $T_j$  を式 (4) によって算出する。ただし、 $(p \wedge q)_i = \min(p_i, q_i)$ ,  $|p|_i = \sum |p_i|$  である。また、 $\alpha$  は選択パラメータと呼ばれる適当な正のパラメータである。
- $$T_j = \frac{|I(t) \wedge w_j(t)|}{\alpha + |w_j(t)|}, \quad (j = 1, \dots, N) \quad (4)$$
- (3)  $T_j$  が最大となる  $F_2$  層ユニット  $c$  を探索する。
  - (4) 結合荷重  $W_j(t)$  および入力データ  $I$  間の整合度  $A_c$  を式 (5) により算出する。

$$A_c = \frac{|I \wedge W_j(t)|}{|I|}, \quad (j = 1, \dots, N) \quad (5)$$

- (5) 整合度  $A_c$  と分類精度の基準である警戒パラメータ  $\rho \in [0, 1]$  を比較する。
  - (a) 「整合度  $A_c \geq$  警戒パラメータ  $\rho$ 」の場合、入力データ  $I(t)$  に基づき結合荷重を式 (6) により更新する。 $r$  は学習率パラメータである。

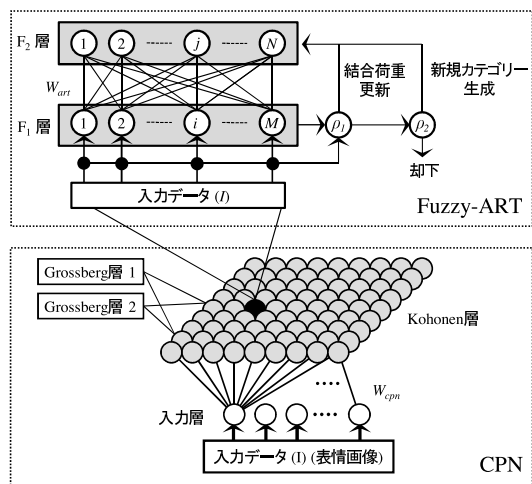


図 5 提案手法  
 Fig. 5 Proposed method.

$$W_j(t+1) = r(I(t) \wedge W_j(t)) + (1-r)W_j(t) \quad (6)$$

- (b) 「整合度  $A_c < \text{警戒パラメータ } \rho$ 」の場合、(a) の条件を満たす  $F_2$  層ユニット  $c$  を探索するまで、選択強度と整合度の算出を繰り返す。
- (c)  $F_2$  層のすべてのユニットで「整合度  $A_c < \text{警戒パラメータ } \rho$ 」の場合、 $F_2$  層に新規のユニット  $N+1$  を追加し、結合荷重  $W$  に入力データ  $I$  の値を設定する (式 (7))。

$$W_{N+1} = I \quad (7)$$

### 3.3 提案手法の処理手順

提案手法は CPN を用いて生成する表情特徴空間 (CPN の Kohonen 層) と、特徴空間の各ユニットに連結した Fuzzy-ART から成る。提案手法の構造を図 5 に示す。また、処理手順を以下に示す。

- (1) 2 章で述べた手法により、CPN を用いて初期学習データを学習し、表情特徴空間を生成する (CPN による初期学習)。
- (2) 表情特徴空間の各ユニット (CPN の Kohonen 層) に Fuzzy-ART を連結し、各ユニットの結合荷重  $W_{cpn}$  を Fuzzy-ART の初期カテゴリの結合荷重  $W_{art}$  として設定する。
- (3) 追加学習データを表情特徴空間に入力し、結合荷重  $W_{cpn}$  とのユークリッド距離が最小となる勝者ユニットを探索する (CPN による表情認識)。
- (4) 勝者ユニット上の Fuzzy-ART が有する既存カテゴリと入力した追加学習データの整合度を算出し、警戒パラメータによる閾値判定の結果から、追加学習処理を実施する (Fuzzy-ART による追加学習)。
- (5) 上記 (3) および (4) の処理を一定期間繰り返した後、表情特徴空間が有する表情カテゴリ (既存の知識、

CPN の結合荷重  $W_{cpn}$ ) と、Fuzzy-ART が有する各カテゴリ (追加学習によって得られた新規の知識、結合荷重  $W_{art}$ ) を再学習データとして設定する。  
 (6) (1) と同様に、CPN を用いた再学習により表情特徴空間を更新する。

上述の (2) から (6) の処理を繰り返すことにより、表情特徴空間は表情認識処理と並行して、既存の知識 (過去に学習した表情パターン) を保持したまま、新たな知識 (未学習の表情パターン) を追加的に学習することが可能となる。

### 3.4 Fuzzy-ART の改良

Fuzzy-ART は入力データと既存カテゴリ間の整合度  $A_c$  を算出し、その値が警戒パラメータ  $\rho$  よりも高い場合は既存カテゴリの結合荷重を更新し、低い場合は新規カテゴリの生成を行う。すなわち、既存カテゴリと著しく異なる表情画像が入力された場合も新規カテゴリが生成される (たとえば、喜びのカテゴリに怒りの表情画像が入力された場合)。このことは、Fuzzy-ART の  $F_2$  層に冗長な知識が増加し、表情特徴空間に誤った知識を生成する要因となる。

そこで提案手法では、図 5 に示すように、既存カテゴリと類似した入力データのみを追加的に獲得することを目的として、既存カテゴリと著しく異なる入力データを排除するための警戒パラメータを追加し ( $\rho_1 > \rho_2$ )、整合度  $A_c$  に対する閾値判定を 2 回行った。これにより 3.3 節の (4) の処理では以下に示す 3 通りの追加学習処理が行われる。

- (1) 「結合荷重更新 ( $A_c \geq \rho_1$ )」:  
 入力データは既存カテゴリに含まれると判断し、既存カテゴリの結合荷重を更新する (既存知識の更新)。
- (2) 「新規カテゴリ生成 ( $\rho_1 > A_c \geq \rho_2$ )」:  
 入力データは既存カテゴリに含まれないものの、類似したデータであると判断し、入力データを新規カテゴリとして追加する (新たな知識の生成)。
- (3) 「却下 ( $\rho_2 > A_c$ )」:  
 入力データは既存カテゴリと著しく異なるデータと判断し、上記の追加学習 (結合荷重更新、新規カテゴリ生成) は行わない (却下)。

上述の処理により、Fuzzy-ART は既存カテゴリに類似した表情のみを追加的に学習し、保有する知識を徐々に拡張することが可能となる。また、冗長な知識の生成を抑制することが可能となる。

## 4. 表情画像の取得方法

本論文では、Ekman 氏によって提唱された人種、文化、性別などによらない普遍的な感情 (怒り、悲しみ、嫌悪、喜び、驚き、恐れ) の 6 種類 [2], [19] の中で、顔表情の物理的变化が最も豊富かつ顕著に表れると考えられる「喜び」

を対象とし、1名の被験者（日本人20代男性）が意図的に表出した喜び表情と無表情の2表情を用いて適応学習機能の検証を実施した。

被験者の顔は正面向きとし、頭部がフレーム内に含まれる状態でUSBカメラ（Logicool社：QcamOrbit）を用いて表情表出の際の動画像（320×240 pixel, 30 fps）を撮影した。具体的には、無表情と喜び表情を交互に複数回表出する過程を撮影し、静止画像に変換して表情画像を取得した。撮影時の室内環境の照明は、日常一般的と考えられる蛍光灯下（約500～700 lx）とし、顔とUSBカメラとの距離は約0.6メートルとしている。また、被験者に対しては表情表出方法に関する指示は行わず、被験者は自身の主観によって喜びの感情に対応する表情の表出を行った。これは、被験者固有の表情表出パターンの取得を目的としたためである。なお、本研究では「秋田大学手形地区におけるヒトを対象とした研究に関する倫理規定第6条第2項」に基づいて倫理審査の申請を行い、承認を得た研究計画の下に被験者本人の了承を得てデータを取得している。

学習に使用する表情画像には顔の構成要素の形状が保持されていることが必要である。本論文では顔の構成要素を含む領域を処理の対象とし、以下の手順に沿って顔領域画像の抽出および正規化を行った。表情画像の正規化手順および結果を図6に示す。

- (1) 照明条件の変動およびノイズ、ならびに画像の解像度（顔領域のサイズ）に対して頑強な Haar-like 特徴 [20] を用いて顔の検出を行い、80×96 pixel のサイズに正規化した顔領域画像を抽出した。
- (2) 表情認識モデルの学習処理を高速化するため、粗視化処理（2×2 pixel のブロック単位）により画像の次元

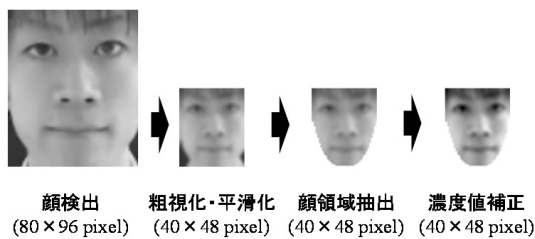


図6 表情画像の正規化

Fig. 6 Normalization process for facial expression image.

数（画像サイズ）を削減した後、平滑化処理を施した（40×48 pixel）。

- (3) 疑似輪郭を生成し、顔の構成要素を含む顔領域を抽出した。具体的には、画像上の (0, 20)–(20, 48) および (40, 20)–(20, 48) の座標間を結ぶ三次曲線を引き、その内部を顔領域として抽出した。
- (4) ヒストグラム線形変換により濃度値補正を行った。

## 5. 追加学習機能に関する基礎実験

### 5.1 実験概要

提案手法の追加学習機能の効果を検証するため、基礎実験を実施した。実験で用いた CPN および Fuzzy-ART の各パラメータの値を表1に示す。使用した表情カテゴリーは喜びと無表情の2表情である。表情特徴空間（CPNのKohonen層）は1次元とし、ユニット数が30, 40, 50ユニットの場合についての比較を行った。実験に使用したデータセットは以下の3種類とし、4章で述べた表情画像から選択して生成した。

- (1) 初期学習データ：100枚  
（口を閉じた喜び表情50枚、無表情50枚）
- (2) 追加学習データ：200枚  
（口を徐々に開く過程の喜び表情）
- (3) 評価データ：2,635枚  
（ランダムに選択した未学習の喜び表情）

追加学習の効果の検証を容易にするため、初期学習データは口を閉じた喜び表情と無表情の2種類で構成し、追加学習データは口を徐々に開く過程の喜びの表情画像を使用した。なお、Fuzzy-ARTの警戒パラメータの値は予備実験の結果から  $\rho_1=0.98$ ,  $\rho_2=0.96$  に設定した。実験手順を以下に示す。

- (1) 初期学習データを用いて初期の表情特徴空間を生成した。
- (2) 追加学習データを表情特徴空間に入力し、追加学習と再学習を繰り返し30回実施することにより、追加学習機能の効果を検証した。
- (3) 30個の表情特徴空間に評価データを入力し、時間軸に対する提案手法の頑強性を評価した。

表1 CPN および Fuzzy-ART のパラメータ

Table 1 Parameters of CPN and Fuzzy-ART.

初期学習 (CPN)	学習データ	100枚 (喜び:50, 無表情:50)
	学習回数	20,000回
	学習率係数	$\alpha=\beta=0.5$
	Kohonen層 (表情特徴空間)	1次元 30, 40, 50 ユニット
追加学習 (Fuzzy-ART)	学習データ	200枚 (喜び)
	警戒パラメータ	$\rho_1=0.98, \rho_2=0.96$
再学習 (CPN)	各種パラメータ	学習データ以外は初期学習と同様
	再学習回数	30回



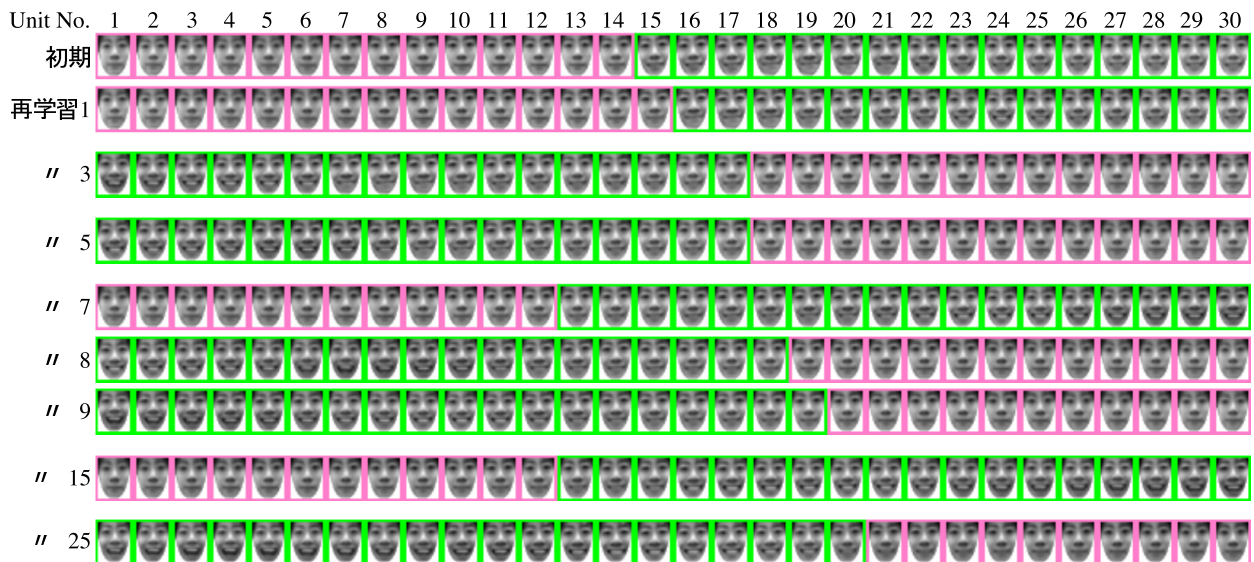


図 7 表情特徴空間における結合荷重の可視化画像 (30 ユニット)

Fig. 7 Visualised images of the weight on the facial expression feature space (30 units).

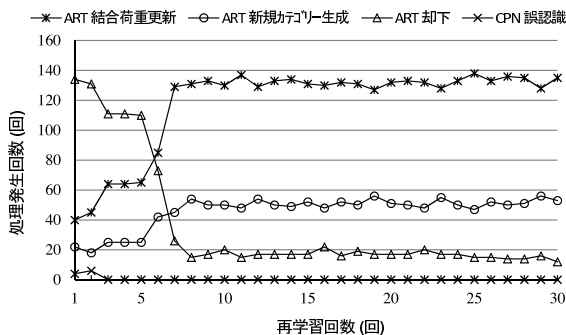


図 8 再学習ごとの追加学習処理発生回数 (50 ユニット)

Fig. 8 Frequency of incremental learning process (50 units).

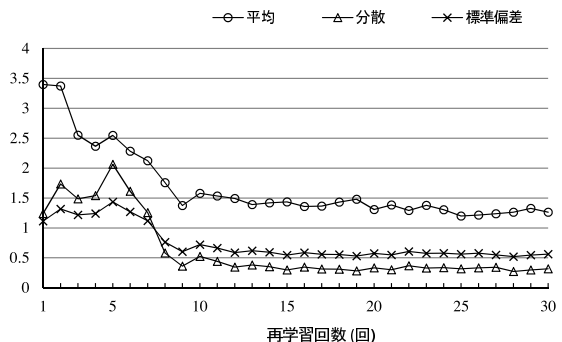


図 9 ユークリッド距離の平均, 分散, 標準偏差 (50 ユニット)

Fig. 9 Average, Variance and standard deviation value of Euclidean distance (50 units).

## 5.2 追加学習機能の検証

### 5.2.1 表情特徴空間の可視化結果に関する考察

生成された表情特徴空間 (30 ユニットの場) の可視化結果 (各ユニットの結合荷重の可視化結果) を図 7 に示す。同図の緑色は喜び表情, 紫色は無表情のカテゴリを表している。同図より, 再学習の増加にともない, 表情特徴空間には徐々に口を開いた喜び表情が獲得されている様子が認められる。一方, 初期の表情特徴空間で学習された口を閉じた喜び表情と無表情の表情パターンは, 再学習を経た後も同様に特徴空間に保持されていることが分かる。上述の内容については, 40, 50 ユニットの表情特徴空間においても同様の傾向が認められた。以上の結果は, 提案手法が既存の知識 (初期に学習した表情パターン) を保持しつつ, 新たな知識 (未学習の表情パターン) を追加的に学習していることを示唆している。

### 5.2.2 追加学習処理の発生回数に関する考察

再学習ごとの追加学習処理 (結合荷重更新, 新規カテゴリ生成, 却下) の発生回数を図 8 に示す。同図は 50 ユニットの表情特徴空間の結果を示している。同図より, 結

合荷重更新, 新規カテゴリ生成処理の発生回数は, 再学習を繰り返すごとに緩やかに増加しており, 再学習回数 7 回目以降で急激に増加し, 収束していることが分かる。一方, 却下処理の発生回数は, 再学習を繰り返すごとに緩やかに減少し, 再学習回数 7 回目以降で急激に減少している様子が認められる。なお, 収束後の数値に着目すると, いずれの再学習においても約 20 回 (20 枚の画像) の却下処理が発生していることが分かる。また, 提案手法は, CPN で入力データと各ユニットの結合荷重のユークリッド距離を算出し, 表情認識を行っている。表情認識の際の再学習ごとのユークリッド距離 (平均, 分散, 標準偏差) の変化を図 9 に示す。同図より, 各値は図 8 と同様に再学習ごとに減少し, 収束していることが分かる。なお, 30, 40 ユニットの表情特徴空間においても同様の傾向が認められた。

再学習を重ねることにより処理の発生回数およびユークリッド距離が収束している要因は, 提案手法が 200 枚の追加学習データの特徴を学習しきったためと考えられる。ここで, 追加学習データに含まれる個々の画像に着目した場

表 2 追加学習処理の時系列変化

Table 2 Time series variation of incremental learning process.

0: ART 結合荷重更新, 1: ART 新規カテゴリー生成, 2: ART 却下, 3: CPN 誤認識

再学習回数	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30			
No.15	1	0	1	0	0	0	0	0	0	0	1	1	0	0	0	1	0	1	0	0	0	1	1	1	1	0	0	0	0	0			
No.69	2	2	0	0	0	0	1	1	0	0	1	1	0	1	0	1	0	1	1	1	0	0	0	1	0	0	0	1	0	0			
No.116	2	2	2	2	2	2	1	1	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	1	0	0	1			
No.159	3	3	2	2	2	2	1	1	1	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	0	0	1	0	0	0			
再学習回数	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30			
No.290	2	2	2	2	3	3	2	2	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	1	2	2	1	2	2		
No.360	3	3	2	2	3	3	3	2	3	2	3	3	3	3	3	3	3	3	3	3	3	3	3	2	1	2	2	2	1	1	2		
No.397	3	3	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3		
No.827	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	2	3	3	3	2	3

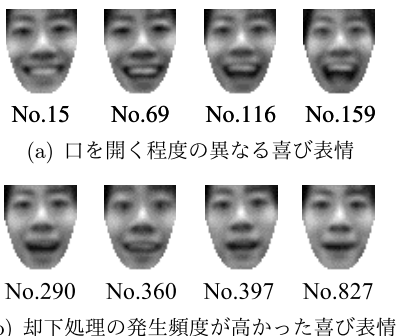


図 10 追加学習データの画像例

Fig. 10 Example of incremental learning data.

合の追加学習処理の変化を表 2 および図 10 示す。

同図 (a) は口を開く程度異なる喜び表情を示しており、(b) は却下処理の発生頻度が高かった喜び表情を示している。表 2 より、同図 (a) の画像に着目すると、口を開く程度が小さい表情から順に新規カテゴリー生成、結合荷重更新処理が行われ、新しい知識として獲得されている様子が認められる。一方、同図 (b) に着目すると、Fuzzy-ART による却下処理、または、CPN による誤認識と判定され、追加学習が行われていないことが分かる。

以上の結果は、提案手法が再学習を重ねるごとに既存カテゴリーに類似した表情のみを追加的に学習し、保有する知識を徐々に拡張していること、また、冗長な知識の生成を抑制していることを示唆している。

### 5.3 未学習データを用いた評価実験

初期学習および追加学習で使用していない未学習の喜びの表情画像 2,635 枚を評価データとして使用し、30 回の再学習で生成された表情特徴空間の評価を実施した。評価に用いた CPN および Fuzzy-ART の各パラメータの値は表 1 の値と同様である。

表情特徴空間に評価データを入力した際の表情認識率を図 11 に示す。提案手法では、入力データが既存の知識と類似している場合に、結合荷重更新および新規カテゴリー生成処理を行っている。したがって、本論文では、結合荷重更新処理と新規カテゴリー生成処理が起こった場合を正しい認識が行われたものとして定義した。同図より、表情特

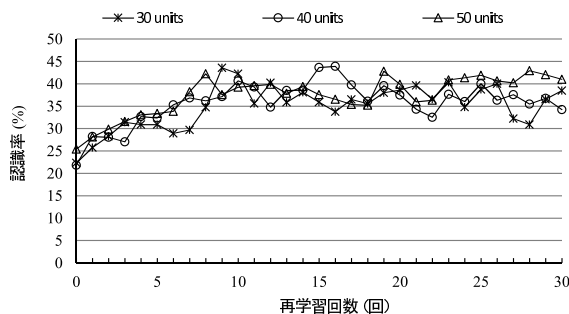


図 11 評価データの表情認識率

Fig. 11 Facial expression recognition rate of evaluation data.

徴空間のサイズが 30 ユニットのときは、初期の表情認識率が 22.4% であるのに対し、再学習 30 回目では 38.5% に増加している。同様に、40 ユニットのときは 21.8% から 34.3% に、50 ユニットのときは 25.4% から 41.0% にそれぞれ増加しており、再学習回数が増加するごとに表情認識率が緩やかに向上していることが分かる。なお、30 ユニットのときの最大認識率は再学習 9 回目の 43.6% であり、40 および 50 ユニットでは、それぞれ再学習 16 回目の 43.9%、28 回目の 42.9% であった。

以上の結果は、再学習回数の増加にともない、表情特徴空間に追加学習データの喜び表情の知識が追加学習され、時間軸に対する汎化能力が向上したことを示唆している。

## 6. 時間経過を考慮した適応学習機能に関する検証

時間経過にともなって出現する未学習の表情パターンを考慮し、11 日間にわたって取得した表情画像を用いて追加学習機能の検証を行った。CPN および Fuzzy-ART のパラメータの値は表 1 の値と同様である。初期学習データは初日に撮影した表情画像 (喜び表情 467 枚、無表情 338 枚の計 805 枚) とし、以降の 10 日分の表情画像を追加学習データとして使用した (650 枚 × 10 日、計 6,500 枚)。なお、追加学習データには喜び表情と無表情の 2 表情を用いた。

初期学習データを用いて初期表情特徴空間を生成し、以降は 1 日分ごとのデータを追加学習、再学習することにより、計 11 個の表情特徴空間を生成した。その後、各特徴空間にすべての追加学習データを評価データとして入力し、



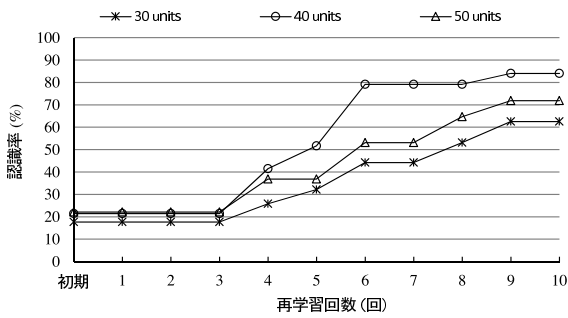


図 12 追加学習データの表情認識率 (6,500 枚)

Fig. 12 Facial expression recognition rate of incremental learning data (6,500 images).

個々の表情認識率を算出した。認識率は、前節と同様に結合荷重更新および新規カテゴリ生成が発生した場合を正しい表情認識が行われたものとして算出した。さらに、初期学習データおよび追加学習データを一度に学習させた表情特徴空間の認識率を算出し（一括学習）、提案手法との比較を行った。

表情認識率の算出結果を図 12 に示す。同図より、30 ユニットの表情特徴空間における提案手法の認識率は初期学習時で 17.7% であるが、再学習を重ねるごとに増加し、9 日目には 62.6% まで向上していることが分かる。一方、一括学習の認識率は 61.0% であったことから、提案手法は一括学習と同等以上の認識性能を有することが明らかとなった。同様に、40 ユニットの表情特徴空間では初期学習時に 21.5% であった認識率が、9 日目には 84.1% まで向上した。50 ユニットのケースも初期学習時の 22.2% から、9 日目には 71.9% まで向上した。両者の一括学習の認識率は、それぞれ 62.7%、70.0% であり、同等以上の認識率となった。以上の結果は、提案手法が未学習の表情パターンを段階的に追加学習していることを示しており、時間経過を考慮した追加学習手法として有用であることを示唆している。

しかし、10 日分の追加学習データの 5 日分については追加学習（結合荷重更新、新規カテゴリ生成処理）が一度も発生せず、すべて却下される場合を認めた（1, 2, 3, 7, 10 日目の追加学習データ）。このことは、学習した知識を保持する表情特徴空間のサイズが小さかったこと、また、追加学習処理の判断基準となる警戒パラメータの値が適切でなかったことに起因すると考える。

## 7. まとめ

本論文では時間軸に対して頑強性を有する表情認識モデルの実現を目的とし、CPN と Fuzzy-ART を応用した適応学習機能を有する表情認識モデルに関して検討を加えた。

本論文で得られた成果を以下にまとめる。

- 追加学習機能の基礎検証として表情特徴空間の結合荷重の可視化画像を解析した結果、提案手法は初期に学習した表情（喜び、無表情）を保持しつつ、喜びの新た

な表情パターンを追加学習していることを確認した。

- 追加学習処理の発生回数に着目した結果、再学習を重ねるごとに結合荷重更新および新規カテゴリ生成の回数は増加し、却下処理の回数は減少すること、また、両者の値は収束することを確認した。さらに、CPN の勝者ユニット探索の際のユークリッド距離の平均、分散および標準偏差に着目した結果、再学習を重ねるごとに各値は減少し、収束することを確認した。
- 11 日間にわたって取得した表情画像を用いて提案手法および一括学習の表情認識率を比較した結果、提案手法は一括学習と同等以上の認識性能を有することを明らかにした。

以上の結果は、提案手法が既存の知識（過去に学習した表情）を保持しつつ、新たな知識（未学習の表情）を段階的に追加学習していることを示している。加えて、提案手法は再学習を重ねるごとに既存カテゴリに類似した表情のみを追加的に学習し、保有する知識を徐々に拡張すること、また、冗長な知識の生成を抑制することを示唆しており、時間経過に対して頑強な表情認識手法として有用であることを示唆している。今後は適切な警戒パラメータの値の設定手法に関する検討、ならびに表情特徴空間を 2 次元に拡張し、6 基本感情を対象とした評価実験を行う予定である。

謝辞 本研究は JSPS 科研費 25700010 の助成を受けた。

## 参考文献

- [1] Birdwhistell, R.L.: Kinesics and Context—Essays on Body Motion Communication, University of Pennsylvania Press (1970).
- [2] Ekman, P.: Darwin and Facial Expression: A Century of Research in Review, Academic Press, N.Y. (1973).
- [3] 赤松 茂：人間とコンピュータによる顔表情の認識 [I]—コミュニケーションにおける表情とコンピュータによるその自動解析, 電子情報通信学会誌, Vol.85, No.9, pp.680–685 (2002).
- [4] 赤松 茂：人間とコンピュータによる顔表情の認識 [II]—コンピュータによる顔表情認識技術 (1): 表情による感情の認識, 電子情報通信学会誌, Vol.85, No.10, pp.766–771 (2002).
- [5] 赤松 茂：人間とコンピュータによる顔表情の認識 [III]—コンピュータによる顔表情認識技術 (2): 顔面の表情動作の認識, 電子情報通信学会誌, Vol.85, No.12, pp.936–941 (2002).
- [6] 赤松 茂：人間とコンピュータによる顔表情の認識 [IV・完]—コンピュータによる顔表情認識の展望: 人間による表情認知に学ぶ, 電子情報通信学会誌, Vol.86, No.1, pp.54–61 (2002).
- [7] Pantic, M. and Rothkrantz, L.J.M.: Expert System for Automatic Analysis of Facial Expression, *Image Vis. Comput.*, Vol.8, No.9, pp.881–905 (2000).
- [8] Pantic, M. and Rothkrantz, L.J.M.: Automatic Analysis of Facial Expressions: The State of the Art, *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol.22, No.12, pp.1424–1445 (2000).
- [9] 北村謙典, 武村紀子, 岩井儀雄, 佐藤宏介: 不快な光環境に誘発される無意識的な表情の認識, 電気学会論文誌 C, Vol.134, No.2, pp.218–224 (2014).

- [10] 松久ひとみ, 橋本 学: Gabor 特徴を用いた顔画像からの微妙な表情変化の推定, 映像情報メディア学会誌, Vol.68, No.6, pp.252–255 (2014).
- [11] 石井雅樹, 山屋孝史, 景山陽一, 高橋 毅, 西田 眞: 適応学習機能を有する表情特徴空間の生成と基礎評価, 映像情報メディア学会誌, Vol.70, No.8, pp.188–190 (2016).
- [12] 石井雅樹, 佐藤和人, 間所洋和, 西田 眞: 自己写像特性を用いた顔表情カテゴリーの抽出と感情空間マップの生成, 電子情報通信学会論文誌, Vol.J91-D, No.11, pp.2659–2672 (2008).
- [13] Ishii, M., Shimodate, T., Kageyama, Y., Takahashi, T. and Nishida, M.: Study on the Emotion Quantification Method using the Facial Expression Feature Space, *International Journal of the Society of Materials Engineering for Resources*, Vol.20, No.1, pp.59–64 (2014).
- [14] 小川行宏, 加藤邦人, 山本和彦: 表情自動学習・認識システムの提案, 電気学会論文誌 C, Vol.124, No.3, pp.699–705 (2004).
- [15] Carpenter, G.A., Grossberg, A. and Rosen, D.B.: Fuzzy-ART: Fast Stable Learning and Categorization of Analog Patterns by an Adaptive Resonance System, *Neural Networks*, Vol.4, pp.565–588 (1991).
- [16] Nielsen, R.H.: Counterpropagation Networks, *Applied Optics*, Vol.26, No.23, pp.4979–4984 (1987).
- [17] Kohonen, T.: Self-organizing Maps, Springer Series in Information Sciences (1995).
- [18] Russell, J.A. and Bullock, M.: Multidimensional Scaling of Emotional Facial Expressions: Similarity from Preschoolers to Adults, *J. Personality and Social Psychology*, Vol.48, pp.1290–1298 (1985).
- [19] Ekman, P. and Friesen, W.V.: Facial Action Coding System, Consulting Psychologist Press (1978).
- [20] Lienhart, R. and Maydt, J.: An Extended Set of Haar-like Features for Rapid Object Detection, *Proc. IEEE Int. Conf. Image Processing*, Vol.1, pp.900–903 (2002).

推薦文

本論文では、時間軸に対して頑強性を有する表情認識モデルの実現を目的とし、対向伝播ネットワークと適応共鳴理論を応用した適応学習機能を有する表情認識モデルを提案している。すなわち、提案手法は、対向伝播ネットワーク (CPN) を用いて生成される表情特徴空間と表情特徴空間の各ユニットに連結した適応共鳴理論 (ART) から構成され、既存の Fuzzy-ART に警戒パラメータをもう 1 つ追加し、既存カテゴリと著しく異なるデータ (喜びのカテゴリに怒りの表情画像が入力された場合など) を排除する機能を追加している。喜びと無表情の 2 表情を対象とした実験の結果、提案手法は (1) 追加的に学習を重ねることによって、一括学習と同等以上の認識率が得られること、(2) 時間経過を考慮した場合の学習においても、学習した知識を保持しつつ、追加学習可能であることを明らかにしており、高く評価できる。このように、提案手法は、知的なヒューマンマシンインターフェイスを実現するための要素技術の 1 つとなるものであり、この分野への大きな寄与が期待される。よって本論文を推薦する。

(東北支部支部長 大堀 淳)



石井 雅樹 (正会員)

1999 年秋田大学大学院鉱山学研究科博士前期課程情報工学専攻修了。同年秋田県工業技術センター情報システム開発部技師。2008 年秋田大学大学院工学資源学研究科博士後期課程電気電子情報システム工学専攻修了。同年秋

田県立大学システム科学技術学部助教、2017 年准教授、現在に至る。画像処理、機械学習を応用したヒューマンマシンインタフェース、知能ロボットに関する研究に従事。博士 (工学)。1998 年素材物性学国際会議優秀論文賞、2003 年情報処理学会東北支部奨励賞、20015 年 IEEE GCCE 2015 Excellent Poster Award (1st prize) 等受賞。電気学会、電子情報通信学会、日本ロボット学会、映像情報メディア学会、画像電子学会等の会員。



山屋 孝史 (正会員)

2014 年秋田大学工学資源学部情報工学科卒業。2016 年同大学大学院工学資源学研究科修了。同年日本ビジネスシステムズ (株) 入社。在学中は主として顔表情の認識に関する研究に従事。



景山 陽一 (正会員)

1997 年秋田大学大学院鉱山学研究科博士前期課程情報工学専攻修了。同年秋田大学助手、2001 年講師、2004 年助教、准教授を経て 2013 年秋田大学大学院工学資源学研究科教授。現在、秋田大学大学院理工学研究科教授。リ

モートセンシングデータの解析技術とアルゴリズムの開発、情景画像における物体認識、ヒューマンセンシングに関する研究に従事。博士 (工学)。1998 年国際素材物性学会議優秀論文賞等受賞。電気学会、電子情報通信学会、映像情報メディア学会、システム制御情報学会等の会員。



高橋 毅 (正会員)

1998年秋田大学鉱山学部電気電子工学科卒業。同年カルソニック(株)入社。2003年秋田大学工学資源学部技官。現在、秋田大学大学院理工学研究科技術専門職員。口唇の色彩情報および形状情報に着目した個人識別・発話認識インタフェースに関する研究に従事。博士(工学)。2011年電気学会優秀論文発表賞, 情報処理学会東北支部奨励賞受賞。電気学会, 日本素材物性学会の会員。



西田 眞 (正会員)

1974年秋田大学鉱山学部電気工学科卒業。同年トヨタ自動車工業(株)入社。1975年秋田大学鉱山学部助手。講師, 助教授を経て, 1996年同教授。2007年総合情報処理センター長, 2010年工学資源学研究科長, 2011年秋田大学理事・副学長, 2014年秋田大学副学長, 2017年放送大学特任教授・秋田学習センター所長, 現在に至る。秋田大学名誉教授。リモートセンシングデータの解析技術とアルゴリズム, 画像情報応用および知識情報システムの研究に従事。工学博士。1988~1989年クラークソン大学客員研究員。1984年電気学会論文賞, 1988年静電気学会論文賞, 1998年国際素材物性学会議優秀論文賞, 2013年日本素材物性学会平成25年度(第23回)山崎賞受賞。電気学会, 映像情報メディア学会, 照明学会, 工学アカデミー, IEEE等の会員。