

# マルチタスク学習用1クラスSVMを用いた 新規人物に対する特異顔表情検知

藤田 裕文<sup>1</sup> 松川 徹<sup>1</sup> 鈴木 英之進<sup>1</sup>

**概要:** 本論文では、複数人物の顔画像データにマルチタスク学習用1クラスSVMを適用し、新規人物の顔画像データより特異顔表情を検知した。この対象問題は異常検知問題の一種であり、不特定者の精神状態推定など種々の重要問題の基盤となる。われわれはまず、多数の無表情顔と幸福顔に、ごく少数の苦痛顔が混在するデータを取得し、顔画像を各局所領域の正規化LBP特徴量に関する主成分で表した。次に1人物を1タスクと見なしてHeらが提案したマルチタスク学習用1クラスSVMを適用した。その結果、1クラスSVMを全タスクおよび各タスクに適用した場合より、平均で0.15高いF値を達成した。

## Detection of Peculiar Facial Expressions to a New Person using Multi-Task Learning with One-Class SVM

HIROFUMI FUJITA<sup>1</sup> TETSU MATSUKAWA<sup>1</sup> EINOSHIN SUZUKI<sup>1</sup>

**Abstract:** In this paper, we applied multi-task learning with one-class support vector machine (OSVM) to image data of multiple persons and detected peculiar facial expressions of a new person. This target problem is a kind of the anomaly detection problem and serves as a basis of various problems including mental state estimation of an unspecified person. We firstly acquired data which include both many images of no-expression or happy faces and few images of painful faces, and represented a facial image with principal components of the normalized LBP features of each local area. Then we regarded one person as one task and applied multi-task learning with OSVM proposed by He et al. As the result, we achieved higher F-measures by 0.15 on average than OSVMs on all tasks and each task.

### 1. 序論

Mehrabianによると、人間のコミュニケーションにおいて、言葉が果たす役割は7%であり、それ以外のノンバーバルな要素の方が重要である[1]。特に顔表情や視線、ジェスチャーなどの視覚情報は、55%を占める[1]、そのうちの1つである顔表情の自動認識の研究が進んでおり、数多くの手法が提案されている[2], [3], [4], [5]。

本研究では、特異顔表情を大多数の顔表情とは特に異なる顔表情であると定義する。特異顔表情は、精神状態の大きな変化に関連すると考えられるため、その検知は重要で

ある。データマイニングにおいて、Knorrらは、大多数のデータとは大きく異なるデータを発見の主目的とする異常検知問題に取り組んだ[6]。顔画像データからの特異顔表情検知は、異常検知問題の一種であり、精神状態推定など種々の重要問題の基盤となる。

手元にデータがまったくない不特定者の特異顔表情検知は、有用であるが、困難であると考えられる。手元のデータが少ない新規人物の特異顔表情検知は、挑戦し甲斐がある現実的な問題である。1人物を1タスクとみなしてマルチタスク学習を用いることにより、手元にデータのある人物の顔表情を手がかりとして、この問題を解決できると期待できる。1クラスSVM(OSVM)は、異常検知手法として長い間用いられてきており、教師なし学習であるためクラスラベル付けの手間が不要であるなどの利点がある[7]。本論文では、異常値検知で使われるOSVMをマルチタスク

<sup>1</sup> 九州大学大学院システム情報科学府, 福岡県福岡市西区元岡 744 番地  
Graduate School of Information Science and Electrical Engineering, Kyushu University, 744, Motoooka, Nishi-ku, Fukuoka, 819-0395, JAPAN

ク用に拡張した手法 [8] を使用する。

本論文の構成は、2 節でマルチタスク学習と表情分析の関連研究を述べ、3 節で顔画像からの特徴量抽出方法と OSVM のマルチタスク学習を説明する。4 節ではデータの収集、実験設定、実験結果を述べ、5 節で結論を述べる。

## 2. 関連研究

### 2.1 マルチタスク学習による顔画像学習

マルチタスク学習では、互いに関連する多数のタスクを同時に学習し、性能の向上を図る [9]。われわれの知る限りでは、マルチタスク学習を用いた特異表情検知は存在しない。ここでは、特異顔表情検知に関連する研究として、顔認識と顔ランドマーク検出にマルチタスク学習を適用し、高正答率を達成した研究を紹介する。

Ding らは、特定の角度から撮影した顔画像データからの顔認識を 1 タスクとして、マルチタスク学習手法を提案した [10]。彼らは辞書学習の一種を用いることにより、顔画像を顔認識に適した空間に写像した。

Masip らは、1 人物を 1 タスクとして、顔認識問題にマルチタスク学習を適用した [11]。1 度学習した後に、新しいタスクが追加された際、前の分類器を拡張する手法を提案した。それにより、すべてのタスクを再学習する必要がなくなった。

Zhang らは、畳み込みニューラルネットワーク (CNN) によるマルチタスク学習を使い、顔のランドマーク検出を行なった [12]。従来の CNN によるマルチタスク学習と比較して、高速に収束する手法を提案した。

論文 [11], [12] では、1 人物を 1 タスクとしてマルチタスク学習を行うことで高い性能を達成している。よって、本論文でも 1 人物を 1 タスクと見なす。

### 2.2 LBP 特徴量を用いた顔表情認識

画像からのパターン認識においては、まず画像データから特徴量を抽出し、その次にパターンを認識する。Local Binary Pattern (LBP) 特徴量 [13] は、種々のパターン認識問題において、高い性能を達成することに貢献してきた。特異顔表情検知問題は、顔表情認識問題の一種である。ここでは、LBP 特徴量を用いた顔表情認識を紹介する。

Luo らは、特徴抽出に LBP と主成分分析 (PCA) を併用し、顔表情を SVM で分類学習した [2]。SVM のみの手法や SVM と PCA を用いた手法と比較して、大幅に高い正答率を達成した。

Moore らは、正面の顔に比較して、顔の角度を変化させた時の顔表情認識精度への影響を調べた [3]。3 次元の顔表情データを扱い、同一表情の様々な角度のデータを作成し、LBP や LBP を拡張した特徴量を用いた。

本論文では、多くの研究で採用されている顔に対する LBP 特徴量抽出方法 [14] を用いた。図 1 右に示すような

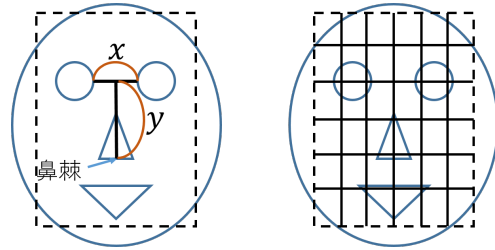


図 1 顔画像の分割

分割した顔ブロック毎に LBP 特徴量を抽出し、それらの LBP 特徴量を連結して顔画像の特徴量とした。

## 3. 特異顔表情検知手法

### 3.1 特徴抽出

画像を入力として、各フレーム画像より LBP 特徴量抽出を行う。本研究で扱う新規人物に対する特異顔表情検知においては、顔の傾きや人物間の顔領域の差異、顔画像間の解像度の差異の 3 つに対処する必要がある。本研究ではそれぞれ、顔微表情検知において用いられた方法 [15] で顔領域分割をした。顔画像の特徴量は、顔領域の補正、LBP 特徴量の抽出の 2 段階で抽出する。

顔領域の補正は、顔の傾きや表情によって不変である左右目頭間の距離と鼻の長さに基づいて行う。図 1 左に示すように、各顔画像において、ライブラリ Dlib[18] を用いて、左右の目頭と鼻棘のそれぞれの特徴点座標を求める。顔の傾きに対処するため、左右の目頭が水平になるように画像を回転させた。さらに、顔領域抽出のための基準となる距離を、左右目頭間の距離  $x$ 、左右目頭の中点と鼻棘との距離  $y$  とした。顔領域は、鼻棘から上方向に  $2y$  の点、下方向に  $y$  の点、左右方向それぞれに  $2x$  の点の 4 点で定義される長方形を顔領域と見なす。

顔画像を図 1 右のように、 $6 \times 6$  のブロックに分割し、各ブロックから抽出した LBP 特徴量 [13] を連結して、ヒストグラムに変換する。ヒストグラムの区間数は、 $256 \times 6 \times 6 = 9216$  である。LBP とは、画像中のそれぞれのピクセルにおいて、その半径  $r$  の範囲に存在する周辺の  $p$  個のピクセルと濃淡値を比べ、周辺のピクセルのほうが高い場合と同等の場合、出力を 1 とし、それ以外を 0 とする。それらの場合の数は  $2^p$  通りある。

$$LBP_{p,r} = \sum_{i=0}^{p-1} s(g_i - g_0) 2^i, \quad (1)$$

$$s(x) = \begin{cases} 1, & x \geq 0, \\ 0, & x < 0. \end{cases} \quad (2)$$

ただし  $g_p$  は周辺のピクセルの濃淡値であり、 $g_0$  は中心の濃淡値である。

高解像度の画像は、ピクセル数が多いため、LBP 特徴量

の性質上、ヒストグラム間における各区間の値は大きい。逆に、低解像度の画像では小さい。顔画像間の解像度の差異の補正は、ヒストグラム度数和の割合に基づく正規化で行う。各顔ブロック  $i$  の LBP ヒストグラム区間  $j$  を  $\mathbf{h}_{i,j}$ 、特微量の次元を  $d$  とすると、正規化後のヒストグラム  $\mathbf{h}'_{i,j}$  は下式になる。

$$\mathbf{h}'_{i,j} = \frac{\mathbf{h}_{i,j}}{\sum_{k=1}^d \mathbf{h}_{i,k}}. \quad (3)$$

ブロックで連結した LBP ヒストグラムに主成分分析を適用し、次元を圧縮する。

### 3.2 OSVM

ここでは、本研究の実験で用いた手法であるマルチタスク用 1 クラス SVM の基となる従来の OSVM を紹介をする。OSVM[7] は、クラスなしデータからの異常検知手法である。すべての例を特徴空間に写像して、原点と大部分の例を最大マージンで分離することにより、異常例の集合を検知する。

$l$  個の学習データ  $\mathbf{x}_1, \dots, \mathbf{x}_l$  を考える。ただし、 $\mathbf{x}_i$  は、空間  $\mathcal{R}^d$  における点である。 $\phi$  を元のデータ空間の点を高次元特徴空間へと写像する関数とする。 $\mathbf{x}$  が正常例か異常例であるかを決定する決定関数  $f(\mathbf{x})$  は、次式で与えられる。 $f(\mathbf{x})$  の符号が負の場合、 $\mathbf{x}$  は異常例であり、それ以外の場合は正常例である。

$$f(\mathbf{x}) = \langle \mathbf{w}, \phi(\mathbf{x}) \rangle - \rho. \quad (4)$$

ここで  $\mathbf{w}$  は、法線ベクトルであり、 $\mathbf{w}$  と  $\rho$  により境界面が決まる。以下の最適化問題を解くことで、大部分の例と原点とを最大マージンで分離することができる。

$$\begin{cases} \min_{\mathbf{w}, \xi, \rho} \frac{1}{2} \|\mathbf{w}\|^2 + \frac{1}{\nu m} \sum_i \xi_i - \rho, \\ \text{s.t. } \langle \mathbf{w}, \phi(\mathbf{x}_i) \rangle \geq \rho - \xi_i, \xi_i \geq 0. \end{cases} \quad (5)$$

ここで、パラメータ  $\nu \in (0, 1)$  は、これは全学習例の内の異常値の割合の上限であり、全例の内のサポートベクトルの割合の下限である。 $\xi_i$  はスラック変数であり、式 (5) の不等式に示すように、最適化の制約を緩和する。式 (5) をラグランジュの未定乗数法で解くことで、パラメータ値  $\mathbf{w}, \xi, \rho$  を決定する。この決定方法は、2 クラス分類学習用のサポートベクトルマシンと類似する [19]。

### 3.3 OSVM のマルチタスク学習

He らは、OSVM をマルチタスク用に拡張した [8]。この手法では、すべてのタスクモデルがある平均モデルに近いとき、各タスクモデル  $f_t$  は一般モデル  $g_0$  と各タスクに特有のモデル  $g_t$  の和で表されると仮定する。すなわち、

$$f_t = g_0 + g_t, \quad (6)$$

ここで、各モデルに式 (4) と同様のモデルを仮定すると下

式になる。

$$f_t = \langle \mathbf{w}_0, \phi_0(\mathbf{x}) \rangle + \langle \mathbf{v}_t, \phi_t(\mathbf{x}) \rangle - \rho_t. \quad (7)$$

ここで、 $\mathbf{w}_0$  と  $\phi_0$  はそれぞれ法線ベクトルと非線形写像関数であり、 $\mathbf{v}_t$  と  $\phi_t$  は、モデル  $t$  の法線ベクトルと非線形写像関数である。この仮定に従い、OSVM をマルチタスク学習に一般化し、以下の最適化問題を得る。

$$\begin{cases} \min_{\mathbf{w}_0, \mathbf{v}_t, \xi_{it}, \rho_t} \frac{1}{2} \sum_{t=1}^T \|\mathbf{v}_t\|^2 + \frac{\mu}{2} \|\mathbf{w}_0\|^2 \\ + \sum_{t=1}^T \left( \frac{1}{\nu_t m_t} \sum_{i=1}^{m_t} \xi_{it} \right) - \sum_{t=1}^T \rho_t, \\ \text{s.t. } \langle \mathbf{w}_0, \phi_0(\mathbf{x}_{it}) \rangle + \langle \mathbf{v}_t, \phi_t(\mathbf{x}_{it}) \rangle \geq \rho_t - \xi_{it}, \xi_{it} \geq 0. \end{cases} \quad (8)$$

ここで、 $T$  はタスク数、 $m_t$  はタスク  $t$  の例数、 $\nu_t$  はパラメータ、 $\xi_{it}$  はタスク  $t$  の  $i$  番目の例のスラック変数、 $\mu$  は一般モデルと各タスクのモデルの類似度を表すパラメータである。各決定関数は、 $\mu$  の値が大きい場合、式 (8) より解の  $\|\mathbf{w}_0\|^2$  が小さくなり、式 (7) より  $f_t$  の符号は主に  $\mathbf{v}_t$  で決まる。よって  $f_t$  は、各タスクそれぞれで学習したモデルに近づく。逆に小さい場合、 $f_t$  はすべてのタスクを一括で学習したモデルに近づく。

式 (8) を、ラグランジュの未定乗数法で解き、双対問題を得る。

$$\begin{cases} \max_{\alpha_{it}} -\frac{1}{2} \sum_{t=1}^T \sum_{r=1}^T \sum_{i=1}^{m_t} \sum_{j=1}^{m_r} \alpha_{it} \alpha_{jr} \\ \left( \frac{1}{\mu} \langle \phi_0(\mathbf{x}_{it}), \phi_0(\mathbf{x}_{jr}) \rangle + \delta_{rt} \langle \phi_t(\mathbf{x}_{it}), \phi_t(\mathbf{x}_{jr}) \rangle \right), \\ \text{s.t. } 0 \leq \alpha_{it} \leq \frac{1}{\nu_t m_t}, \sum_{i=1}^{m_t} \alpha_{it} = 1. \end{cases} \quad (9)$$

$$\delta_{rt} = \begin{cases} 1, & r = t, \\ 0, & r \neq t. \end{cases} \quad (10)$$

一般モデルのカーネルを  $k_0$ 、各タスクのモデルのカーネルを  $k_t$  とすると、特徴空間における例同士の内積を、以下のようになっている。

$$k_0(\mathbf{x}_{it}, \mathbf{x}_{jr}) = \langle \phi_0(\mathbf{x}_{it}), \phi_0(\mathbf{x}_{jr}) \rangle, \quad (11)$$

$$k_t(\mathbf{x}_{it}, \mathbf{x}_{jr}) = \langle \phi_t(\mathbf{x}_{it}), \phi_t(\mathbf{x}_{jr}) \rangle. \quad (12)$$

カーネルの特性を利用し、以下の新しいカーネル  $G_{rt}$  を定義する。

$$G_{rt}(\mathbf{x}_{it}, \mathbf{x}_{jr}) = \frac{1}{\mu} k_0(\mathbf{x}_{it}, \mathbf{x}_{jr}) + \delta_{rt} k_t(\mathbf{x}_{it}, \mathbf{x}_{jr}). \quad (13)$$

このカーネルを用い、以下のように決定関数を表す。

$$f_t(\mathbf{x}) = \text{sign} \left( \sum_{r=1}^T \sum_{i=1}^{m_r} \alpha_{ir} G_{rt}(\mathbf{x}_{ir}, \mathbf{x}) - \rho_t \right). \quad (14)$$

表 1 データセットのそれぞれの表情の枚数

|       | 学習データ |      | テストデータ |      |      |
|-------|-------|------|--------|------|------|
|       | 無表情   | 幸福表情 | 無表情    | 幸福表情 | 苦痛表情 |
| 参加者 1 | 410   | 490  | 11     | 5    | 8    |
| 参加者 2 | 581   | 319  | 10     | 6    | 8    |
| 参加者 3 | 492   | 408  | 11     | 5    | 8    |

## 4. 実験

### 4.1 データ収集

本実験では、特異顔表情として苦痛顔の表情を選択した。3 名の実験参加者に無表情、幸福顔、苦痛顔を意識的に作ってもらい、カメラ正面を向いた状態をビデオカメラで撮影した。幸福顔と苦痛顔を撮影する際は、それらのモデル画像をみせ、その画像の表情を真似するように指示した。画像の解像度は  $1920 \times 1080$  である。各タスクでの無表情、幸福顔、苦痛顔の枚数は表 1 の通りである。

### 4.2 実験設定

特異表情を正クラス、その他を負クラスとする。テストデータの中の正クラス例の数を  $N_P$ 、正クラス例で正クラスと予測された例数を  $N_{TP}$ 、負クラス例で正クラスと予測された例数を  $N_{FP}$  とする。再現率  $R$ 、適合率  $P$ 、F 値  $F$  を以下のように定義する。

$$R = \frac{N_{TP}}{N_P}, \quad (15)$$

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}}, \quad (16)$$

$$F = \frac{2RP}{R + P}. \quad (17)$$

2 名の学習データを用いて、残りの 1 名のテストデータに対して予測する。その 1 名の学習データのうち、無表情 3 枚、幸福顔 3 枚を学習データに加えた。テストデータには、その 6 枚のデータは含まれていない。

比較手法は、各タスクで独立して学習する T-OSVM と全タスクを一括で学習する 1-OSVM である。T-OSVM では、予測対象タスクの無表情 3 枚、幸福顔 3 枚を学習データとして用いて、1-OSVM では、それら 6 枚と他タスクの全学習データを学習データとして用いた。本実験では、式 (8) のパラメータ  $\nu$  を 0.1 とし、式 (1) のパラメータ  $p$  を 8,  $r$  を 1 とした。

表 3 OCSVM の適用結果

|       | 1-OSVM |      |      | T-OSVM |      |      |
|-------|--------|------|------|--------|------|------|
|       | 再現率    | 適合率  | F 値  | 再現率    | 適合率  | F 値  |
| タスク 1 | 1      | 0.33 | 0.50 | 1      | 0.36 | 0.53 |
| タスク 2 | 0      | 0    | N/A  | 1      | 0.33 | 0.50 |
| タスク 3 | 1      | 0.4  | 0.57 | 1      | 0.33 | 0.50 |

### 4.3 実験結果

式 (8) のパラメータ  $\mu$  を変化させ、一般モデルと各タスクのモデルの類似度を変化させた。He らの手法での再現率、適合率を表 2、比較手法の再現率、適合率を表 3 にそれぞれ示す。N/A は計算不可を表す。

表 3 を見ると、1-OSVM と T-OSVM 共に、再現率が 1、適合率が 0.33 になっている場合が多い。これは、テストデータの全てを異常値と判定していることを表し、異常検知手法としては不適切である。1-OSVM をタスク 2 に適用した場合、すべての正クラス例を負クラスと予測し、再現率、適合率が 0 になった。

図 2, 3, 4 は、各タスクを新規タスクにした時の結果である。黒い線は境界線であり、内側が異常値と予測される。 $\mu$  が小さいと、他のタスクの影響を受け、境界線が広がっていることが分かり、 $\mu$  が大きいと、対象タスクの 6 つの学習例に合わせた境界線になっていることがわかる。対象タスクの学習データは少数であるが、 $\mu$  を調整することで一般モデルの知識を利用しつつ、対象タスクに特有のモデルを学習できていることがわかる。

## 5. 結論

精神状態推定など種々の重要問題の基盤となる特異表情検知を行った。予測対象は、少数のデータしか得られない新規人物であった。顔画像の顔をブロックに分割し、LBP ヒストグラムを抽出し、それを特徴量とした。さらに、He らが提案したマルチタスク学習用の 1 クラス SVM を適用した。その結果、通常の 1 クラス SVM とシングルタスク学習用 1 クラス SVM と比較して、平均で 0.15 高い F 値を達成することに成功した。

式 (8) のパラメータ  $\mu$  を変化させ、境界面の変化を可視化し、一般モデルとタスクに特有のモデルを組み合わせることで、境界面の内側に少数の例のみが存在することを防ぐことを確認した。

本論文では、特異顔を苦痛顔として、新規人物の少数の

表 2  $\mu$  を変化させた時の再現率と適合率

|       | $\mu = 0.001$ |      |      | $\mu = 1$ |      |      | $\mu = 1000$ |      |      | $\mu = 10000$ |      |      |
|-------|---------------|------|------|-----------|------|------|--------------|------|------|---------------|------|------|
|       | 再現率           | 適合率  | F 値  | 再現率       | 適合率  | F 値  | 再現率          | 適合率  | F 値  | 再現率           | 適合率  | F 値  |
| タスク 1 | 1             | 0.53 | 0.69 | 1         | 0.53 | 0.69 | 1            | 0.50 | 0.67 | 1             | 0.50 | 0.67 |
| タスク 2 | 1             | 0.61 | 0.76 | 1         | 0.61 | 0.76 | 1            | 0.57 | 0.73 | 1             | 0.57 | 0.76 |
| タスク 3 | 0.25          | 0.25 | 0.25 | 0.38      | 0.30 | 0.34 | 0.88         | 0.41 | 0.56 | 0.88          | 0.41 | 0.56 |

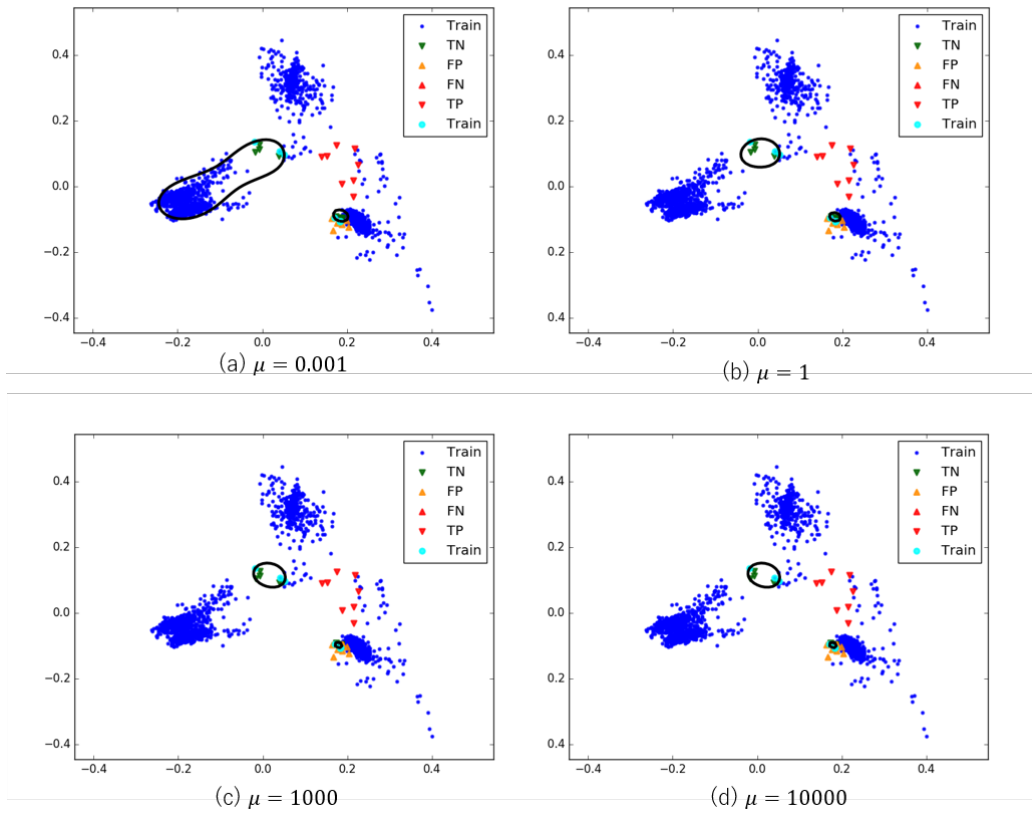


図 2  $\mu$  を変化した時のタスク 1 に対しての実験結果

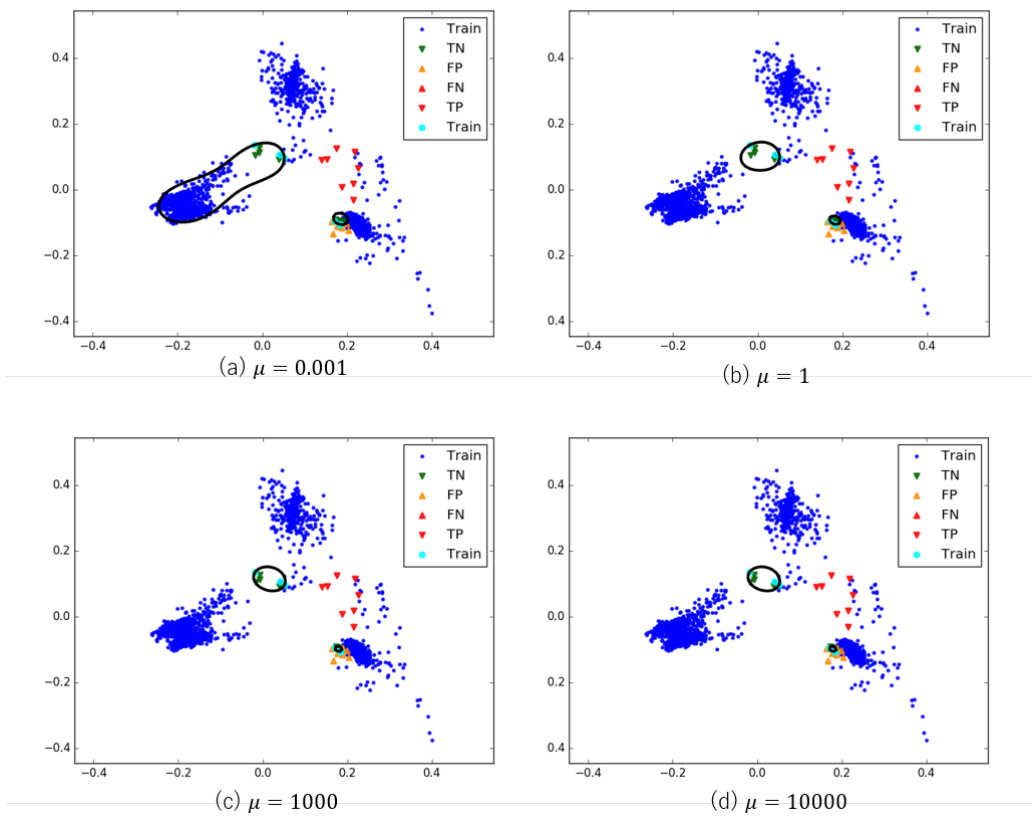


図 3  $\mu$  を変化した時のタスク 2 に対しての実験結果

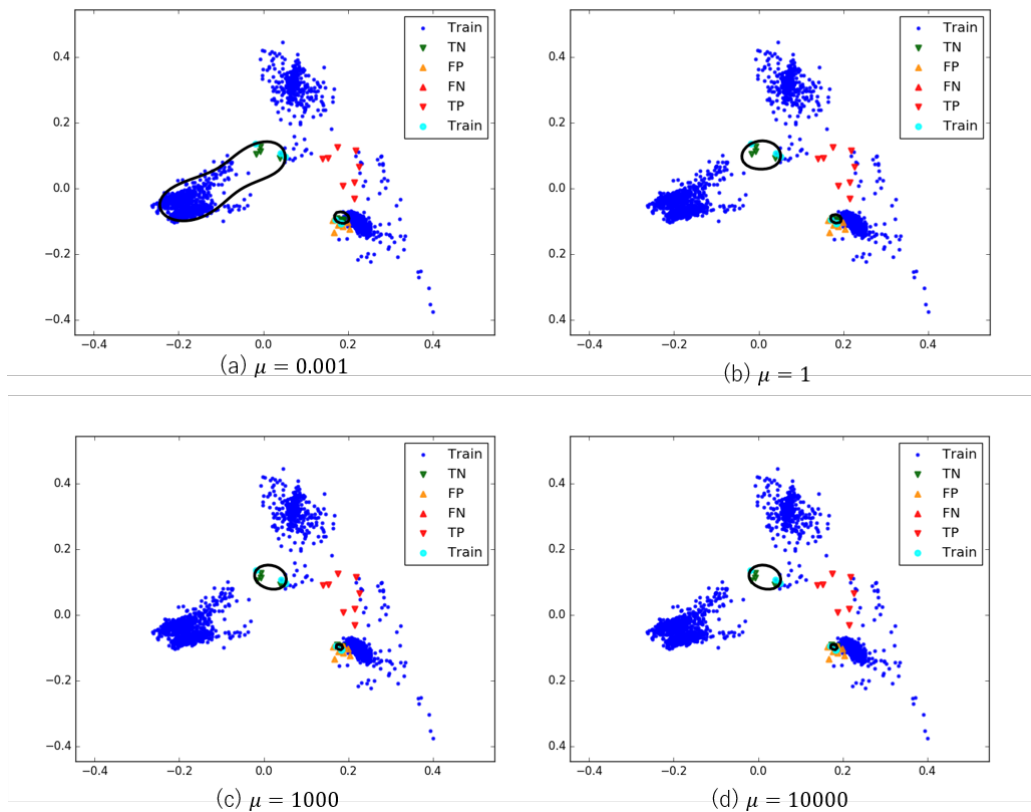


図 4  $\mu$  を変化した時のタスク 3 に対しての実験結果

非特異顔データを使用し学習した。次の課題は、特異顔を苦痛顔として、新規人物のデータをまったく使わない実験設定下で適切な境界面を設定することである。

## 謝辞

本研究のデータ収集に協力された 2 名に感謝する。本研究の一部は JSPS 科研費 15K12100 の助成を受けている。

## 参考文献

- [1] Mehrabian, A.: Verbal and Nonverbal Interaction of Strangers in a Waiting Situation., *Journal of Experimental Research in Personality* (1971).
- [2] Luo, Y., Wu, C.-m. and Zhang, Y.: Facial Expression Recognition Based on Fusion Feature of PCA and LBP with SVM, *Optik-International Journal for Light and Electron Optics*, Vol. 124, No. 17, pp. 2767–2770 (2013).
- [3] Moore, S. and Bowden, R.: Local Binary Patterns for Multi-View Facial Expression Recognition, *Computer Vision and Image Understanding*, Vol. 115, No. 4, pp. 541–558 (2011).
- [4] Abdulrahman, M., Gwadabe, T. R., Abdu, F. J. and Eleyan, A.: Gabor Wavelet Transform Based Facial Expression Recognition Using PCA and LBP, *Signal Processing and Communications Applications Conference (SIU), 2014 22nd*, IEEE, pp. 2265–2268 (2014).
- [5] Shan, C., Gong, S. and McOwan, P. W.: Facial Expression Recognition Based on Local Binary Patterns: A Comprehensive Study, *Image and Vision Computing*, Vol. 27, No. 6, pp. 803–816 (2009).
- [6] Knorr Edwin, M. and Ng Raymond, T.: Algorithms for Mining Distance-Based Outliers in Large Datasets [R], *Proceedings of the 24th International Conference on Very Large Data Bases*. New York: MorganKaufmann, pp. 392–403 (1998).
- [7] Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J. and Williamson, R. C.: Estimating the Support of a High-Dimensional Distribution, *Neural Computation*, Vol. 13, No. 7, pp. 1443–1471 (2001).
- [8] He, X., Mourot, G., Maquin, D., Ragot, J., Beausery, P., Smolarz, A. and Grall-Maës, E.: Multi-Task Learning with One-class SVM, *Neurocomputing*, Vol. 133, pp. 416–426 (2014).
- [9] Caruana, R.: Multitask learning, *Machine Learning*, Vol. 28, No. 1, Springer, pp. 41–75 (1997).
- [10] Ding, C., Xu, C. and Tao, D.: Multi-Task Pose-Invariant Face Recognition, *IEEE Transactions on Image Processing*, Vol. 24, No. 3, pp. 980–993 (2015).
- [11] Masip, D., Lapedriza, A. and Vitrià, J.: Multitask Learning-An Application to Incremental Face Recognition., *VISAPP (1)*, pp. 585–590 (2008).
- [12] Zhang, Z., Luo, P., Loy, C. C. and Tang, X.: Facial Landmark Detection by Deep Multi-Task Learning, *European Conference on Computer Vision*, pp. 94–108 (2014).
- [13] Ojala, T., Pietikäinen, M. and Harwood, D.: A Comparative Study of Texture Measures with Classification Based on Featured Distributions, *Pattern Recognition*, Vol. 29, No. 1, pp. 51–59 (1996).
- [14] Ahonen, T., Hadid, A. and Pietikainen, M.: Face Description with Local Binary Patterns: Application to Face Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 12, pp. 2037–2041 (2006).
- [15] Moilanen, A., Zhao, G. and Pietikäinen, M.: Spotting

- Rapid Facial Movements from Videos Using Appearance-Based Feature Difference Analysis, *IEEE 22nd International Conference on Pattern Recognition (ICPR)*, pp. 1722–1727 (2014).
- [16] Liao, S., Zhu, X., Lei, Z., Zhang, L. and Li, S. Z.: Learning Multi-Scale Slock Local Binary Patterns for Face Recognition, *International Conference on Biometrics*, pp. 828–837 (2007).
- [17] Ojala, T., Pietikainen, M. and Maenpaa, T.: Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 7, pp. 971–987 (2002).
- [18] Kazemi, V. and Sullivan, J.: One Millisecond Face Alignment with an Ensemble of Regression Trees, *2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1867–1874 (2014).
- [19] Suykens, J. A. and Vandewalle, J.: Least Squares Support Vector Machine Classifiers, *Neural Processing Letters*, Vol. 9, No. 3, pp. 293–300 (1999).