顔特徴点を用いた特徴選択と特徴抽出による表情認識に基づく映像中の表情表出シーン検出

野宮　浩揮†　　宝珍　輝尚†

† 京都工芸繊維大学大学院工芸科学研究科
〒 606–8585 京都府京都市左京区松ヶ崎御所海道町
E-mail: †{nomiya,hochin}@kit.ac.jp

あらまし　印象的な映像の検索に用いることを目的として，映像中の人物の表情を認識し，様々な表情が現れているシーンを検出する手法を提案する．表情表出時における眉や目の端点等の顔特徴点の位置変化を用いて，2 つの顔特徴点間の距離および 3 つの顔特徴点で囲まれる領域の面積や領域内の画素の輝度に基づいて，多様な顔特徴量を定義する．顔特徴量の中には冗長なものもあるので，顔特徴量の有用性を定義し，有用な特徴量のみを選択して表情認識に用いる．さらに，選択された特徴に対する特徴抽出により得られる特徴ベクトルを用いて機械学習を行うことで，効率的かつ正確な表情認識モデルの構築を図る．また，映像中の各フレームに対する表情認識の結果から表情表出の開始点と終了点を検出し，表情が表出されているシーンを特定する．

キーワード　表情認識，特徴選択，主成分分析，AdaBoost，マルチメディアデータベース

Emotional Video Scene Detection based on Facial Expression Recognition by Feature Selection and Extraction using Facial Feature Points

Hiroki NOMIYA† and Teruhisa HOCHIN†

† Graduate School of Science and Technology, Kyoto Institute of Technology
Goshokaido-cho, Matsugasaki, Sakyo-ku, Kyoto 606–8585 Japan
E-mail: †{nomiya,hochin}@kit.ac.jp

1. はじめに

近年における画像や映像の記録機器の高性能化や，画像・映像データを大量に保存することのできる大容量ストレージの出現により，例えば旅行や年中行事などの様子をビデオカメラで撮影したものなど，個人でも多くのマルチメディアデータを所有する機会が増大している．また，ライフログのように，個人の行動や体験の履歴をマルチメディアデータとして記録し，記憶の補助やパターン分析などに利用するといったことも行われている [1] [2].

しかし，その一方で，記録するデータの量が膨大になってしまったため，画像や映像などの検索が難しくなり，記録したマルチメディアデータが十分に利用されなくなる問題が発生している．そのため，大量のマルチメディアデータから，的確かつ効率的に検索を行いたいという要求が高まってきている．そこで本論文では，マルチメディアデータの中でも，特に効率的な検索が要求される映像データから，検索の対象となりやすいと考えられる印象的なシーンを検出する手法について検討する．印象的なシーンには様々なものがあるが，その中には映像中の人物の感情が変動し，それに伴って表情が変化するものが多く存在すると考えられるので，本論文では，映像中の人物に様々な表情が表出しているシーンを検出することを目的とする．

表情表出シーンを検出するためには，表出されている表情を的確に識別することのできる表情認識手法が不可欠である．しかし，人の表出する表情は様々であり，また個人差も大きいことから，多様な表情を正確に見分けることは難しいという問題がある．表情ごとに人手で識別モデルを構成すれば，正確な識別が可能になる反面，専門的知識や多くの人的コストが要求される [3] [4]．また，表情表出時には顔面上に複雑な動きが生じるため，適切に顔面上の変化を捉え，記述することが難しいという問題もある．顔の三次元モデルなど，複雑な特徴量を扱うことのできる識別モデルを構成すれば，表情表出時の顔面上の微妙な変化を捉えることができ，正確な表情認識が可能になるが，識別モデルが複雑になるため，モデル構築が難しく，また

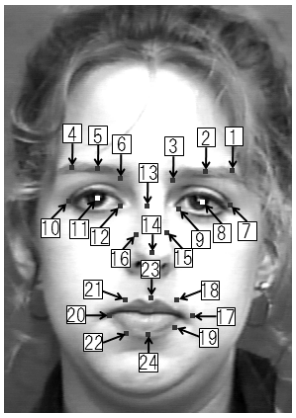

図 1　顔 特 徴 点

認識に要する時間が長くなると考えられる [5] [6].

そこで，これらの問題を解決するため，簡潔な特徴量を多数組み合わせて学習を行うことにより，表情認識時の計算コストを抑えつつ，認識性能の高い表情認識モデルの構築を図る．表情表出時には，眉と目の端点や中心点，あるいは鼻と口の周辺の点などの特定の点（これ以降，顔特徴点と呼ぶ）が特徴的に動くと考えられるので，顔特徴点の動きに着目して，表情認識に用いる特徴量を定める．顔特徴点は特徴量を簡潔に表現できるため，表情認識によく用いられるが [7] [8]，多くの認識手法では，予め定められた特徴点のみを用いるため，特定の表情は正確に識別できても，多様な表情の識別に適用することは難しいと考えられる．そこで，提案手法では，様々な顔特徴点の組み合わせから多数の特徴量を定義し，学習用の顔画像を用いて，認識対象とする表情に応じて識別に有効な特徴量を選択することで，多様な表情の識別に適用できる認識手法を構成する．また，主成分分析に基づく特徴抽出を行うことにより，認識精度を低下させることなく，より簡潔に特徴量の内容を表現できる特徴ベクトルを生成し，表情認識の効率化を図る．

映像から表情表出部を検出するため，映像中の各フレームの画像に対して前述の表情認識を行い，表情表出区間と表出している表情を推定する．フレームによっては，顔特徴点抽出の誤りや顔面の動きによる画像のぶれ，映像中の人物の瞬きなどの影響で適切な認識ができないことがあるので，そのようなフレームによる検出性能の低下を抑えるための表情表出部検出法を構成する．また，表情表出の様子を捉えた映像データを用いた表情表出部検出実験により，提案手法の検出性能を評価する．

以降，2 章では提案手法で用いる顔特徴点および特徴量を示し，3 章で表情表出部の検出法について述べる．4 章で実験を通じて提案手法の有効性を評価し，5 章でまとめを行う．

2.　認識に用いる特徴量

2.1　顔 特 徴 点

顔特徴点は，眉，目，鼻，口など表情の表出に深く関与すると考えられる顔面の構成要素の位置や形状の変化が反映されるように定める [9]．具体的には，図 1 に示す 24 個の点を顔特徴点として用いる．

2.2　顔 特 徴 量

表情が表出される際には，複数の顔特徴点の位置が同時に変化すると考えられるため，複数の顔特徴点の相互関係に基づいて，以下の特徴量を定める．

- 直線特徴 [9]
  無表情時と表情表出時における，2 つの顔特徴点間の距離をそれぞれ L, L' としたときの距離の比 L'/L．任意の 2 つの特徴点について，合計 276 種類定義される．
- 三角形特徴
  3 つの顔特徴点を結んでできる三角形領域に基づいて以下の 3 種類を定める．いずれも任意の 3 つの特徴点について，合計 2024 種類定義される．
  - 輝度平均による三角形特徴 [9]
    無表情時と表情表出時における，三角形領域内の画素の輝度の平均値をそれぞれ I, I' としたときの平均値の比 I'/I.
  - 輝度ヒストグラムによる三角形特徴 [9]
    三角形領域内の画素の輝度ヒストグラム（4 つの階級をもつ）から得られる．無表情時と表情表出時における，i 番目の階級値を H_i, H'_i とすると，$\frac{H'_i}{P'} - \frac{H_i}{P}$ $(i = 1, 2, 3, 4)$ を要素にもつ 4 次元実数値ベクトルで表される．なお，P, P' はそれぞれ無表情時と表情表出時の三角形領域に含まれる画素数である．
  - 面積による三角形特徴
    無表情時と表情表出時における，三角形領域の面積をそれぞれ A, A' としたときの面積比 A'/A.

これらの特徴量は，画像中の顔の大きさ（スケール）の違いや，顔の傾きなどによる特徴量への影響を低減するために，いずれも顔特徴点の相対的な位置関係に基づいて定めている．

2.3　画像のフィルタリング

顔の構成要素の輪郭（エッジ）は照明条件の変化などに対して頑健であると考えられるため，顔画像に対して，次の 2 種類のフィルタを用いてエッジを抽出して得られた画像に対しても特徴量を求める．

- 水平 Prewitt フィルタ
  水平方向のエッジを検出できる 1 次微分フィルタ.
- 垂直 Prewitt フィルタ
  垂直方向のエッジを検出できる 1 次微分フィルタ.

与えられた顔画像（原画像）に加えて，これらのフィルタを適用した画像からも輝度平均および輝度ヒストグラムによる三角形特徴を求める．したがって，特徴量の総数は $276 + 6072 + 6072 + 2024 = 14444$ 種類となる．

3.　表情認識手法

3.1　学習に用いる事例

表情認識モデルを構築するために，予め表出している表情が与えられている事例を訓練事例として用いて学習を行う．訓練事例の数を n とすると，訓練集合は $\{(x_1, x'_1, y_1), \ldots, (x_n, x'_n, y_n)\}$ と表される．ここで，x_k は訓練事例中の k 番目の事例の無表情時の顔画像から得られる顔特徴点の集合であり

($x_k = \{p_1, \ldots, p_{24}\}$)，$x'_k$ は表情表出時の顔画像から得られる顔特徴点の集合である ($x'_k = \{p'_1, \ldots, p'_{24}\}$)．また，$y_k$ は表出している表情であり，認識対象とする表情の数を C 種類とすると，$y_k \in \{1, \ldots, C\}$ となる．すなわち，各事例は無表情時と表情表出時の顔画像から得られる顔特徴点の集合および表出している表情の三つ組により構成される．

3.2 確率密度分布に基づく表情識別関数

各特徴量の分布は正規分布に近くなると考えられるので，正規分布の確率密度関数に基づいて，与えられた事例がどの表情を表しているかを推定する識別関数 N_{ci} を各表情 c および各特徴量 f_i ($i = 1, \ldots, 14444$) について以下のように定める．

$$N_{ci}(x, x') = \frac{1}{\sqrt{2\pi}\sigma_{ci}} \exp\left\{-\frac{(f_i(x, x') - \mu_{ci})^2}{2\sigma_{ci}^2}\right\}. \quad (1)$$

ここで，$f_i(x, x')$ ($i = 1, \ldots, 14444$) は無表情時の顔画像 x と表情表出時の顔画像 x' から求められる，i 番目の特徴量の値を示しており，$f_1, \ldots, f_{276}$ が直線特徴，$f_{277}, \ldots, f_{6348}$ が輝度平均による三角形特徴，$f_{6349}, \ldots, f_{12420}$ が輝度ヒストグラムによる三角形特徴，$f_{12421}, \ldots, f_{14444}$ が面積による三角形特徴に対応している．なお，輝度平均と輝度ヒストグラムによる三角形特徴の f_i の値は，ベクトルの各要素について求める．また，μ_{ci} と σ_{ci}^2 はそれぞれ表情 c が表出している事例の特徴量 f_i の平均と分散であり，以下の式を用いて求められる．

$$\mu_{ci} = \frac{1}{n_c} \sum_{k=1}^{n} \left\{I(y_k = c) \cdot f_i(x_k, x'_k)\right\}, \quad (2)$$

$$\sigma_{ci}^2 = \frac{1}{n_c - 1} \sum_{k=1}^{n} \left\{I(y_k = c) \cdot (\mu_{ci} - f_i(x_k, x'_k))^2\right\}. \quad (3)$$

ここで，$I(\alpha)$ は α が成立するときに 1 となり，成立しないときに 0 となる関数である．また，n_c は表情 c が表出している事例の総数であり，$n_c = \sum_{k=1}^{n} I(y_k = c)$ と表される．

3.3 有用な特徴量の推定

特徴量をすべて用いると計算量の増大が問題となるため，表情の識別に有効な特徴量のみを選択的に用いることで表情認識を効率化する．特徴量の有用度は，級間分散と級内分散の比に基づいて定める．まず，特徴量 f_i の級間分散 B_i は次のように表される．

$$B_i = \frac{\sum_{c=1}^{C} n_c(\nu_i - \mu_{ci})^2}{C - 1}. \quad (4)$$

ここで，ν_i は全クラスの特徴量の平均値を平均した値であり，以下の式を用いて求められる．

$$\nu_i = \frac{1}{C} \sum_{c=1}^{C} \mu_{ci}. \quad (5)$$

また，f_i の級内分散 W_i は次のように表される．

$$W_i = \frac{\sum_{c=1}^{C} \sum_{k=1}^{n} \left\{I(y_k = c) \cdot (f_i(x_k, x'_k) - \mu_{ci})^2\right\}}{n - C}. \quad (6)$$

f_i の有用度 U_i は，級間分散と級内分散の比 B_i/W_i と定める．

すべての特徴量に対して有用度を求め，有用度の高い特徴量から順にいくつかの特徴量を用いて認識を行う．なお，認識に使用する特徴量数は, [9] で用いている手法により決定する．

3.4 表情表出モデルの構築

前述の特徴量はいずれも単純であり，高速に求められるが認識精度は不十分であると考えられる．そこで，単純な識別器を統合することで識別性能を向上することのできる AdaBoost に基づくアンサンブル学習を行う [10]．また，3.3 節で述べた特徴量の有用性算出法に基づくと，相関の高い特徴量が複数選択されることがあるため，選択された特徴量をすべて用いると冗長性が高くなる可能性がある．この問題を回避するため，選択された特徴量から構成される特徴ベクトルに対して，主成分分析に基づく特徴抽出を行い，累積寄与率が 90% 以上となる最小の数の主成分得点の値を並べて生成される特徴ベクトルを学習に用いる．図 2 に，アンサンブル学習アルゴリズムを示す．

（1） 訓練集合中の全事例の重みを $w_i = \frac{1}{n}$ ($i = 1, \ldots, n$) とする．

（2） For $t = 1$ to T (T はラウンド数)

(a) 重み w_i にしたがって訓練事例のランダムサンプリングを行い，第 t ラウンドでの訓練集合 X_t を構成する．

(b) X_t を用いて Support Vector Machine により弱学習器 $h_t(x)$ を生成する．この際，主成分分析を行った後の特徴ベクトルを用いる．

(c) h_t の誤り率 ε_t を式 (7) により求める ($x_i \in X$).

$$\varepsilon_t = \frac{\sum_{i=1}^{n} w_i \cdot I(h_t(x_i) \neq y_i)}{\sum_{i=1}^{n} w_i}. \quad (7)$$

(d) h_t の有用性 α_t を式 (8) により求める．

$$\alpha_t = \log \frac{1 - \varepsilon_t}{\varepsilon_t} + \log(C - 1). \quad (8)$$

(e) X 中の各事例の重み w_i を式 (9) にしたがって更新する．

$$w_i \leftarrow w_i \cdot \exp\{\alpha_t \cdot I(h_t(x_i) \neq y_i)\} \quad (i = 1, \ldots, n). \quad (9)$$

(f) X 中の各事例の重み w_i を式 (10) にしたがって正規化する．

$$w_i \leftarrow \frac{w_i}{\sum_{j=1}^{n} w_j}. \quad (10)$$

（3） 各ラウンドで生成された弱学習器を統合し，最終的な識別モデル $H(x)$ を以下のように生成する．

$$H(x) = \operatorname*{argmax}_c \sum_{t=1}^{T} \alpha_t \cdot I(h_t(x) = c). \quad (11)$$

図 2　アンサンブル学習アルゴリズム

3.5 瞬きの影響の低減

瞬きをしているフレームでは目が閉じられているため，驚きの表情など表情表出中に目が開いていることの多い表情の表出部を検出する際に，誤識別の要因となる可能性がある．そこで，瞬きをしていると判断されたフレームについては，それらのフレームに対する表情認識結果を用いるのではなく，そのフレームの周囲の瞬きをしていないと判断されたフレームの表情認識結果を用いて表情の推定を行う．

3.5.1 瞬きの判定

瞳の部分が見えていれば瞬きをしていないと考えられるので，瞳がどの程度見えているかを瞬き判定の基準とする．瞳の部分

を検出しやすくするため，瞬きの判定には，図 3 に示すように，左右の目の周囲の領域から得られる垂直エッジ画像を用いる．なお，図 3 に赤枠で示しているこの領域は，抽出された顔特徴点の位置を用いて定める．左目（右目）の領域は，図 1 の 7 番と 9 番（10 番と 12 番）の顔特徴点を左右の端点とし，横幅の半分の長さを縦幅とする．また，8 番（11 番）の顔特徴点が上端と下端の中間の位置となるように定める．

図 3　瞬き判定に使用する領域（赤枠内の領域）

図 3 の左右の目の領域画像それぞれに対して，式 (12) に示す 7×7 画素のカーネルフィルタ K を適用する．カーネルフィルタとは，画像内のある一定の領域に含まれる画素に対して特定の演算を行う行列であり，画像からエッジや特定の形状を抽出する際に用いられる．

$$K = [K_{i,j}] = \begin{pmatrix} -2 & -1 & 0 & 2 & 0 & -1 & -2 \\ -1 & 0 & 2 & 0 & 2 & 0 & -1 \\ 0 & 2 & 0 & -1 & 0 & 2 & 0 \\ 2 & 0 & -1 & -4 & -1 & 0 & 2 \\ 0 & 2 & 0 & -1 & 0 & 2 & 0 \\ -1 & 0 & 2 & 0 & 2 & 0 & -1 \\ -2 & -1 & 0 & 2 & 0 & -1 & -2 \end{pmatrix}. \quad (12)$$

このカーネルフィルタを画像の (x, y) の位置にある画素（この画素の輝度値を $P_{x,y}$ とする）に適用すると，以下の式 (13) で表される値 $\kappa_{x,y}$ が得られる．

$$\kappa_{x,y} = \sum_{i=1}^{7}\sum_{j=1}^{7} K_{i,j} P_{x+j-4,y+i-4}. \quad (13)$$

このカーネルフィルタでは，正の係数が円形に近い形で配置されているため，円形に近いエッジがある部分では κ の値が大きくなる傾向がある．瞳の部分はほぼ円形で個人差も小さいので，瞳の見えている程度が κ の値に比例すると考えられる．

目を開いている場合，κ の値は瞳の付近で高い値となり，それ以外の領域では低い値となると考えられる．また，目を閉じている場合は，全体的に低い値をとると考えられる．そこで，両目の画像の各画素から得られる κ の不偏分散値 V を用いて瞬きの判定を行う．V は以下の式 (14) で表される．

$$V = \frac{1}{2}\left\{ \frac{\sum_{y=4}^{H^L-3}\sum_{x=4}^{W^L-3}(\kappa^L_{x,y} - \overline{\kappa^L})^2}{(W^L-6)(H^L-6)-1} + \frac{\sum_{y=4}^{H^R-3}\sum_{x=4}^{W^R-3}(\kappa^R_{x,y} - \overline{\kappa^R})^2}{(W^R-6)(H^R-6)-1} \right\}. \quad (14)$$

ここで，W^L, H^L（W^R, H^R）はそれぞれ左目（右目）の領域画像の横幅と縦幅を表し，$\overline{\kappa^L}$（$\overline{\kappa^R}$）は $\kappa^L_{x,y}$（$\kappa^R_{x,y}$）の平均値を表す．なお，抽出される特徴点の位置にはノイズが混入することがあるため，各フレーム画像について，前後 2 フレームを含めた 5 つのフレームでの V の平均値を用いることにより，ノイズの低減を図る．

瞬きをすると，V の値が急激に低下した直後に急激に上昇する傾向が見られる．そこで，i 番目のフレームでの V の値 V_i について，前後 M フレームの区間での $V_{i-M}, V_{i-M+1}, \ldots, V_i, \ldots, V_{i+M}$ の各値が式 (15) と式 (16) の条件をみたしていれば，i 番目のフレームで瞬きが行われていると判断する．

$$V_{i-1} \geqq V_i \ \wedge \ V_i \leqq V_{i+1}. \quad (15)$$

$$\frac{\max\{V_{i-M},\ V_{i-M+1},\ \ldots,\ V_{i-1}\}}{V_i} \geqq \theta \ \wedge \ \frac{\max\{V_{i+1},\ V_{i+2},\ \ldots,\ V_{i+M}\}}{V_i} \geqq \theta. \quad (16)$$

ここで，M と θ は閾値である．式 (15) が成立することは，V_i の値が極小になっていることを意味する．また，式 (16) が成立することは，i 番目のフレームの前 M フレームおよび後ろ M フレームの中に，それぞれ V の値が V_i の θ 倍以上になるフレームが少なくとも 1 つ以上存在することを意味する．

3.5.2　瞬きをしているフレームに対する表情推定

瞬きをしていると判断されたフレームに対しては，式 (11) から得られる表情認識結果を用いず，次に述べる手法で表出している表情を推定する．瞬きをしていると判断された i 番目のフレームの表情 $\hat{E}_i$ は以下の式 (17) で推定する．なお，瞬き開始フレームを i_s 番目のフレーム，瞬き終了フレームを i_e 番目のフレームとし，$i_s \leqq i \leqq i_e$ とする．

$$\hat{E}_i = \underset{e}{\operatorname{argmax}} \left\{ \sum_{j=1}^{m} I(E_{i_s-j} = e) + \sum_{k=1}^{n} I(E_{i_e+k} = e) \right\}, \quad (17)$$

$$m = i_e - i + 1,\ n = i - i_s + 1. \quad (18)$$

ここで，E_j（$j < i_s \vee j > i_e$）は，瞬きをしていないと判断された j 番目のフレームに対する，式 (11) より得られる表情認識結果を表す．したがって，i 番目のフレームの周囲の $(m+n)$ 個の瞬きをしていないと判断されたフレームに対する表情認識結果のうち，最も多い表情が i 番目のフレームで表出していると推定される．

3.6　表情表出区間の検出

映像から表情表出区間を検出するため，すべてのフレームの画像に対して，瞬きをしていないフレームについては 3.4 節で述べた表情認識モデルによる表情認識を行い，瞬きをしているフレームについては 3.5 節で述べた手法による表情認識を行う．そして，得られた結果に基づいて表情表出開始フレームと表情表出終了フレームを定めることで，表情表出区間を決定する．

3.6.1　表情表出開始フレームの検出

次の条件をみたす p 番目のフレームを表情表出開始フレームと定める．

- 第 $(p-N)$ フレームから第 $(p+N)$ フレームまでの間に，表情が表出している（無表情でない）フレームが $(N+1)$ 個以上存在する．
- 第 p' フレームから第 $(p-1)$ フレームまでの間に表情表出開始フレームがない．ここで，p' は最初のフレームまたは直近の表情表出終了フレームの番号であり，$p' < p$ である．

ここで，$N = 3$ のときの表情表出開始フレーム検出の例を図 4 に示す．図のグラフの横軸がフレーム番号，縦軸が各フレームに対する表情認識結果を示している．この場合，前述の条件をみたす p は 11 であるため，第 11 フレームが怒りの表情表出開始フレームとなる．

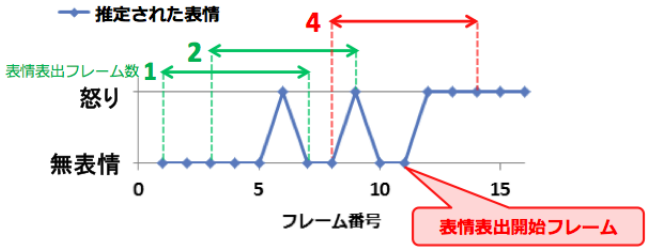

図 4　表情表出開始フレーム検出の例

3.6.2　表情表出終了フレームの検出

次の条件をみたす q 番目のフレームを表情表出終了フレームと定める．

- 第 $(q-N)$ フレームから第 $(q+N)$ フレームまでの間に，表情が表出していない（無表情の）フレームが $(N+1)$ 個以上存在する．
- 第 q' フレームから第 $(q-1)$ フレームまでの間に表情表出終了フレームがない．ここで，q' は直近の表情表出開始フレームの番号であり，$q' < q$ である．

$N = 3$ のときの表情表出終了フレーム検出の例を図 5 に示す．この場合，前述の条件をみたす q は 68 であるため，第 68 フレームが表情表出終了フレームとなる．

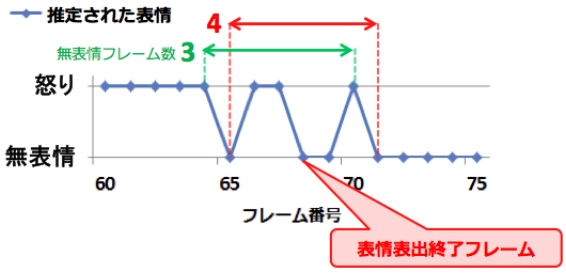

図 5　表情表出終了フレーム検出の例

3.6.3　表情表出区間内で表出されている表情の推定

表情表出区間内の全フレームで投票を行い，最も得票数の多い表情，すなわち最も多くのフレームで表出されている表情をその区間内で表出されている表情と定める．したがって，表情表出区間内で表出されている表情 E は式 (19) で求められる．

$$E = \underset{e}{\operatorname{argmax}} \sum_{n=p}^{q} I(E_n = e). \tag{19}$$

ここで，E_n は 第 n フレームで表出されていると推定された表情である．

4.　実　　験

4.1　使用したデータセット

提案手法の有効性を評価するため，MMI Facial Expression Database [12]（以下 MMI データセットと表記する）を用いて表情認識実験および表情表出区間検出実験を行った．なお，MMI データセットには，18〜63 歳の男女の被験者 29 人の表情データが，表情が表出して無表情に戻るまでの様子を映した動画として収められている．動画のフレームレートは 25 フレーム／秒である．この中から，15 人の被験者による計 100 個の事例を用いた．

本実験では，認識対象の表情は怒り，嫌悪，恐怖，幸福，悲しみ，驚き，無表情の 7 種類としている．各表情の事例数はそれぞれ 14, 13, 11, 28, 13, 21 である．また，訓練集合とテスト集合に同一人物の事例（動画）が含まれないようにして，10 分割の交差検定を行った．

4.2　静止画像に対する表情認識実験

はじめに，提案手法の表情認識性能を評価するため，各動画の無表情時および表情表出時の静止画像のみを用いて，表情認識実験を行った．無表情時の静止画像として動画の最初のフレーム画像を用い，表情表出時の静止画像として動画の中間地点でのフレーム画像を用いた．実験的に，AdaBoost のラウンド数を 50，3.6.1, 3.6.2 節で述べた N の値を 3 と定めた．また，無表情に対する認識精度を評価するため，無表情の事例を 56 個作成し，認識対象とした．

4.2.1　認 識 精 度

まず，提案手法の認識精度を各表情に対して評価する．表 1 に各表情に対する再現率，適合率と F 値およびデータセット全体の認識精度を示す．

表 1　各表情に対する再現率，適合率，F 値およびデータセット全体の認識精度

表情	再現率	適合率	F 値
無表情	0.857	0.686	0.762
怒り	0.357	0.625	0.455
嫌悪	0.692	0.750	0.720
恐怖	0.167	0.200	0.182
幸福	0.821	0.697	0.754
悲しみ	0.077	0.500	0.133
驚き	0.571	0.571	0.571
認識精度	63.46%		

表 1 より，無表情と幸福，嫌悪，驚きの表情は比較的正確に認識できる傾向にある．これは，例えば幸福の表情では表情表出時に口角が上がる，驚きの表情では目が大きく見開かれる，嫌悪の表情では眉間に皺が寄る，無表情では顔特徴点の位置が

ほとんど変化しないなど，他の表情にはあまり見られない特徴があるため，他の表情と見分けやすいからであると考えられる．一方，恐怖や悲しみの表情は比較的誤識別が多い．悲しみの表情については，表情表出時に顔面上の変化が小さく，無表情と判断される場合があることが影響していると考えられる．恐怖の表情は，表情表出時に驚きや悲しみの表情と類似した動き（目が見開かれる，眉が下がるなど）が見られることがあるため，これらの表情と混同されることが原因であると考えられる．

4.2.2 認識速度

次に，提案手法の効率性を評価するため，認識速度の計測を行った．認識速度の計測に用いた計算機に搭載されている CPU は Xeon W3580 (3.33GHz)，メモリは 8GB である．なお，マルチスレッド処理は行っていない．

およそ 40 万画素から構成される事例（顔画像）を用いて表情認識を行った結果，1 つの事例について，認識に使用する特徴量の抽出処理（顔特徴点の抽出処理は除く）に要した平均時間が約 0.41 秒，表情認識モデルによる表情の認識処理に要した平均時間が約 0.09 秒であった．このことから，ある程度短い時間で認識を行えるといえる．また，認識に要する時間の大半が画像からの特徴量抽出処理によるものであり，この処理に要する時間は概ね画像の大きさに比例することから，画像のサイズを認識に支障がない程度に縮小することにより，認識精度を低下させずに認識速度を向上できる可能性があると考えられる．

4.3 表情表出部検出実験

MMI データセットに含まれる各動画の全てのフレーム画像を用いて，表情表出部検出実験を行った．実験的に，3.5.1 節で述べた M の値を 7，θ の値を 1.5 と定めた．その他の設定は 4.2 節での実験と同一である．表 2 に各表情に対する検出精度およびデータセット全体の検出精度を示す．なお，表 2 には，動画中の全フレームの 80% 以上のフレームの表情を正しく認識できた場合を正解とした際の検出精度を示している．これは，検出結果を著者の一人が目視で確認したところ，全体の 80% 以上のフレームの表情を正しく認識できれば，表情が表出している部分を概ね検出できる傾向があることを確認したためである．

表 2　各表情に対する検出精度およびデータセット全体の検出精度

表情	検出精度 (%)
怒り	14.29
嫌悪	53.85
恐怖	18.18
幸福	85.71
悲しみ	7.69
驚き	66.67
全体	50.00

本実験においても，静止画像に対する表情認識実験と同様に，嫌悪，幸福，驚きの表情については，比較的高い検出精度が得られた．これらの表情は他の表情と比べて認識しやすく，また無表情とも区別しやすいため，動画中の多くのフレーム画像に対して正しい表情認識ができたことが主な要因と考えられる．

一方，恐怖，悲しみの表情の検出精度は低くなっている．これも，静止画像に対する表情認識にも見られる傾向であるため，各フレームでの認識精度が十分でないことが表情表出区間の検出性能に影響を与えていると考えられる．怒りの表情については，静止画像での認識精度はそれほど低くないが，表情表出部検出精度は低くなっている．これは，表情が強く表れているフレーム画像の表情は正しく認識されたものの，表情表出開始・終了時付近のフレームの多くが無表情であると判断され，怒りの表情が表出していると判断された区間の長さが不十分になったことで不正解となる傾向が見られたことが原因である．

ここで，幸福の表情が表出しているある事例に対する表情表出部検出結果の一例を示す．まず，瞬きの検出を行う前のフレーム毎の表情認識結果（各フレーム画像に対する，式 (11) から得られた表情認識結果）および瞬きと判断されたフレームを図 6 に示す．図に示しているグラフは，横軸がフレーム番号，縦軸が各フレームに対する表情認識結果を表している．

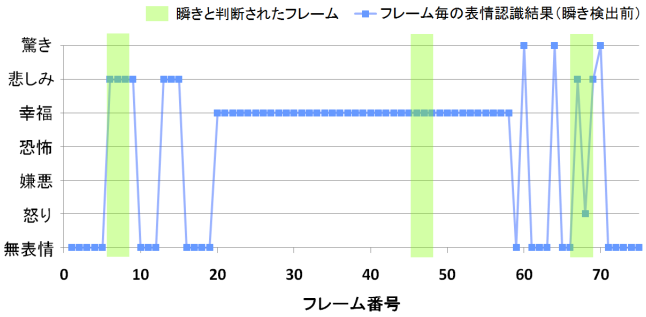

図 6　瞬き検出前のフレーム毎の表情認識結果と瞬きと判断されたフレーム

図 6 より，瞬きの検出を行う前のフレーム毎の表情認識結果では，表情表出中は概ね正しく表情を認識できているが，表情表出前および表出終了後に表情を誤識別しているフレームがいくつか見られる．特に，瞬きをしていると判断されているフレームに誤識別が多い．瞬きをしていると判断されている区間は全部で 3 つあり，これらの区間の中央のフレーム画像は図 7 のようになっている．なお，これら 3 つの区間以外に，映像中の人物が実際に瞬きをしている区間はない．

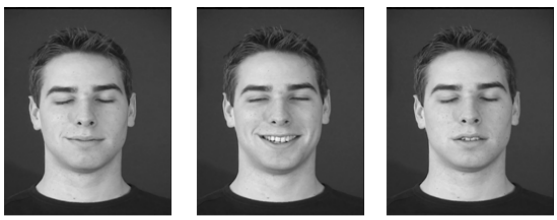

図 7　瞬きをしていると判断されている区間の中央のフレーム画像

図 7 より，2 番目の区間（46〜48 番目のフレーム）では口の形状から幸福の表情であると判断できるため，瞬きをしていても正しく表情が認識できたものと考えられる．一方，1, 3 番目の区間（6〜8 番目，67〜69 番目のフレーム）は本来無表情で

あるが，多くのフレームが悲しみの表情に誤識別されている．これは，悲しみの表情には目を伏せているものが多く，瞬きにより目を閉じている状態を，悲しみにより目を伏せている状態であると誤って判断した結果であると考えられる．

このように，瞬きをしているフレームでは誤識別が多く見られるが，3.5 節で述べた瞬きをしているフレームに対する表情推定法を用いることによって，1, 3 番目の区間のすべてのフレームが無表情であると適切に判断された．また，2 番目の区間についても，すべてのフレームが幸福の表情であるという推定結果が得られた．その結果，最終的な表情表出部検出結果は図 8 のようになり，この推定結果は正解と判断された．比較のため，図 8 には正解と定めた表情表出部も示している．

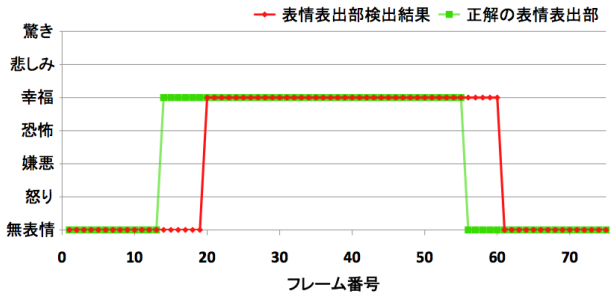

図 8　表情表出部検出結果と正解の表情表出部

この結果から，検出された幸福の表情表出区間は正解の表情表出区間に近くなっていることがわかる．両者の区間に多少のずれはあるものの，ずれはおよそ 5 フレーム程度であるので，時間にして約 0.2 秒と非常に短い．また，図 9(a), (b) に示すように，両者の表情表出開始フレームおよび表情表出終了フレームの画像には大きい違いはないといえる．

このように，幸福や驚きの表情については，表情表出中のフレームでの認識が概ね正しいことと，無表情のフレームで誤識別が生じた場合でも，3.5 節と 3.6 節で述べた手法の導入により，誤りを訂正できる可能性があることから，比較的高い検出精度が得られたといえる．

次に，誤検出が非常に多かった悲しみの表情が表出しているある事例に対する表情表出部検出結果の一例を示す．図 10 に瞬き検出前の各フレーム毎の表情認識結果と正解の表情表出部を示す．

図 10 より，一部のフレームは悲しみの表情であると正しく認識されているが，多くのフレームが無表情であると誤認識されていることがわかる．その結果，全体として無表情であると判断され，表情表出部が検出されないという結果になっている．ここで，この事例における主要なフレーム画像を図 11 に示す．図 11(a) は無表情のフレーム画像の一例であり，最初のフレームの画像である．(b), (d) はそれぞれ正解の表情表出開始・終了フレームの画像である．(c) は悲しみの表情が強く表れているフレームの一例であり，提案手法により悲しみの表情であると判断されたフレームの画像である．

図 10 より，悲しみの表情が強く表れているフレーム画像においても，無表情のフレーム画像と比べて口と眉の周囲が多少

図 9　検出された表情表出区間と正解の表情表出区間における表情表出開始・終了フレーム画像および表情が強く表れているフレーム画像の一例

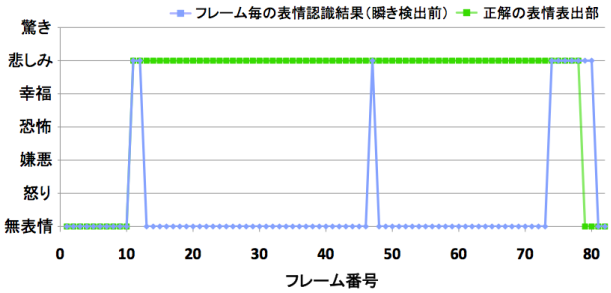

図 10　瞬き検出前の各フレーム毎の表情認識結果と正解の表情表出部

変化しているだけであり，変化に乏しいといえる．この傾向は多くの悲しみの表情の事例に見られるため，悲しみの表情の検出精度が低い要因であると考えられる．また，この傾向は怒りや恐怖の表情の事例の一部にも見られるため，これらの表情の検出精度が低い一因であるといえる．しかし，これらの表情では，多くのフレームで（無表情ではない）他の表情に誤識別されることも検出精度を低下させる原因となっているため，検出性能の向上のためには，各フレーム画像に対する表情認識をより正確に行う必要がある．

4.4　有用と判断された特徴量

最後に，特徴量の有用性の平均値の順位を表 3 に示す．なお，輝度平均と輝度ヒストグラムによる三角形特徴には，適用したフィルタの種類を「原画像」（フィルタ不適用），「水平」（水平 Prewitt フィルタ），「垂直」（垂直 Prewitt フィルタ）の形式で併記している．また，輝度平均・輝度ヒストグラム・面積による三角形特徴は，それぞれ「輝度平均」，「輝度ヒストグラム」，

図 11　主要なフレーム画像

「面積」と略記している．

表 3　特徴量の有用性の平均値の順位

順位	特徴量
1	直線
2	輝度平均（垂直）
3	輝度ヒストグラム（水平）
4	輝度平均（水平）
5	輝度平均（原画像）
6	面積
7	輝度ヒストグラム（垂直）
8	輝度ヒストグラム（原画像）

表 3 より，直線特徴の順位が最も高いことから，表情表出時における，顔特徴点の位置関係が表情の識別に有効であると考えられる．また，輝度ヒストグラムによる三角形特徴よりも輝度平均による三角形特徴の方が有効と判断されやすい傾向にあることから，画素の輝度については平均値を用いる方が効果的であるといえる．面積による三角形特徴は，輝度平均による三角形特徴ほどは有効ではないが，輝度ヒストグラムによる三角形特徴よりは有用となる可能性がある．

顔特徴点の部位の観点から述べると，有効と判断された特徴量としては，眉や目の周囲の顔特徴点と口の周囲の顔特徴点をつなぐ直線特徴や，眉と口の周囲の顔特徴点あるいは目と口の周囲の顔特徴点から構成される，顔の広い範囲を含む三角形特徴が多いという傾向が見られた．

5. ま と め

顔特徴点から様々な特徴量を生成して，その中から有用な特徴量のみを選択し，特徴抽出を行って表情を認識する手法，および映像中の各フレームに対する表情認識結果から，表情表出区間を検出する手法を提案した．評価実験の結果，表情表出時の顔面の動きが大きい表情や，他の表情と大きく異なる表情は比較的正確に認識でき，表情表出部が精度よく検出されることを示した．

しかし，悲しみなど表情表出時に顔面に特徴的な動きが見られない表情は誤認識しやすい傾向にあるため，微妙な表情の違いを識別できる特徴量や認識モデルの構成法を検討する予定である．さらに，幸福と驚きなど，複数の感情が混合されているように見える表情が表出されることもあるので，そのような場合にも対応できる，より正確かつ柔軟な表情認識法の確立を図る．また，映像の照明条件の変化に対する頑健性の検証や，表情の認識に必要な最小限の画像サイズを判断し，特徴量の計算コストを抑えることによる認識速度の向上も今後の課題である．

謝辞　本研究の一部は，科学研究費補助金（課題番号：22700098）による．ここに記して謝意を表します．

文　献

[1] J. Gemmell, G. Bell, R. Luederand, S. Drucker and C. Wong, “MyLifeBits: Fulfilling the Memex Vision,” Proc. of the 10th ACM International Conference on Multimedia, pp. 235–238, 2002.

[2] Y. Kono and K. Misaki, “Remembrance Home: Storage for Re-discovering One’s Life,” Proc. of Pervasive 2004 Workshop on Memory and Sharing of Experiences, pp. 25–30, 2004.

[3] 太田 寛志，佐治 斉，中谷 広正，“顔面筋に基づいた顔構成要素モデルによる表情変化の認識，” 電子情報通信学会論文誌，Vol. J82-D-II, No. 7, pp. 1129–1139, 1999.

[4] I. Essa and A. Pentland: “Coding, Analysis, Interpretation, and Recognition of Facial Expressions,” IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 757–763, 1997.

[5] F. Dornaika and F. Davoine: “Simultaneous Facial Action Tracking and Expression Recognition Using a Particle Filter,” Proc. of IEEE International Conference on Computer Vision, pp. 1733–1738, 2005.

[6] 根岸 秀行，米田 政明，酒井 充，長谷 博行，東海 彰吾，“顔部品に依存しない正面顔を用いた表情認識手法，” 電子情報通信学会技術研究報告 パターン認識・メディア理解，Vol. 104, No. 447, pp. 37–42, 2004.

[7] 小林 宏，丹下 明，原 文雄，“人の顔の 6 基本表情の実時間認識，” 日本ロボット学会誌，Vol. 14, No. 7，pp. 80–88, 1996.

[8] I. Hupont, E. Cerezo and S. Baldassarri, “Sensing Facial Emotions in a Continuous 2D Affective Space,” Proc. of 2010 IEEE International Conference on Systems, Man, and Cybernetics, pp.2045–2051, 2010.

[9] 野宮浩揮，宝珍輝尚，“表情表出時の顔特徴点の変化を用いたアンサンブル学習による表情認識，” 電子情報通信学会技術研究報告 Vol.110, No. 187, PRMU2010-68, IBISML2010-40, pp. 85–92, 2010.

[10] J. Zhu, H. Zou, S. Rosset and T. Hastie, “Multi-class Adaboost,” Statistics and Its Interface, Vol. 2, pp. 349–360, 2005.

[11] T. Kanade, J. F. Cohn and Y. Tian, “Comprehensive Database for Facial Expression Analysis,” Proc. of the 4th IEEE International Conference on Automatic Face and Gesture Recognition, pp. 46–53, 2000.

[12] M. Pantic, M. F. Valstar, R. Rademaker and L. Maat, “Web-based database for facial expression analysis,” Proc. of IEEE International Conference on Multimedia and Expo, pp. 317–321, 2005.