表情認識における基準表情の自動抽出と表情推移の認識

学生員 海老根 秀 之 (工学院大学)

非会員 志 賀 泰 博 (工学院大学)

非会員 池 田 美 鈴 (工学院大学)

正 員 中 村 納 (工学院大学)

Automatic Detection of Reference Face and The Recognition of Transition of Facial Expressions

Hideyuki Ebine, Student Member, Yasuhiro Shiga, Non-member, Misuzu Ikeda, Non-member, Osamu Nakamura, Member (Kogakuin University)

Several researchers have recently attempted automatic recognition and tracking of the facial expressions in video frames. However, it is very hard to detect a neutral face because the distinction between facial movement and head movement is difficult to make. To solve this distinction problem, the head position was initialized using the central point of each eye and the tip of the nose. The transition of facial expressions can be recognized by means of detecting the reference face (neutral face) and the recognition algorithm constructed to the expert system.

キーワード：表情認識, 修正 HSV 表色系, 平常顔検出, 差分画像, 動画像

Keywords: Facial Expressions Recognition, Modified HSV Color System, Neutral Face Detection, Differencial Image, Video Image

1. はじめに

顔表情の認識は、手話の自動通訳、介護補助、マーケティングなど様々な分野への応用が期待されている。

表情は、それが表出した顔である一枚の表情顔から認識することは困難であり、平常顔と比較することで認識できるものと考えられる。例えば、ある表情顔を見たときに怒っている顔であると認識できるかもしれない。実はその顔は、その人にとっては表情が表出していない平常顔であるかもしれない。すなわち、表情は感情刺激により表情筋が緊張・弛緩することによる、顔表面の3次元的な形状変化であると考えられる⁽⁵⁾。従って、この変化を捉えるには、基準となる表情である平常顔からの顔表面の変化を追跡することが必要である。

筆者等は、このような考えに基づき、平常顔と表情顔との差分から表情の認識を行い、ある程度の成果をあげている⁽¹⁾⁻⁽³⁾。しかしながら、さらに精度をあげ、表情の推移までも追跡するためには、平常顔と表情顔の2枚の画像だけでなく、動画像を用いて表情の時系列変化を追跡することが不可欠である⁽⁶⁾⁽¹¹⁾。これには、表情の表出に伴う顔表面の変化を平常顔と比較することにより行うことが必要なため、平常顔のフレームを動画像から抽出しなければならない。しかし、これに関する報告はほとんど成されていない。

平常顔は、一定時間ほとんどその3次元形状が変化しないという特徴を持つ。そこで、筆者等は顔内部の変化量を捉えることで、そのフレームを検出することができるのではないかと考えた。一般に、頭部全体が動いた場合、顔の変化のみを捉えることは極めて困難である。そこで、まず、両目の中心と鼻の頂点を用いて頭の位置を正規化することとした。これにより、顔の変化を捉えることが可能となった。顔部位の変化から平常顔を検出し、検出された平常顔を基準に表情の認識を行う。

2. 認識システム

〈2・1〉 処理手順 本手法は次の5段階から成る。

- (1) 顔領域の抽出
- (2) 頭部の位置の正規化
- (3) 平常顔の検出
- (4) 特徴量の抽出
- (5) 表情の認識

〈2・2〉 顔領域の抽出 入力画像は平行光源の基でCCDカメラによって撮影された、720×486画素、各色256段階のカラー画像である。入力画像の例を図1(a)に示す。

背景や頭髮、衣服といった部分は表情の認識には特に必要ではない。そこで、前処理として、入力画像から背景や髪の毛などを除去し、顔領域のみを抽出する(図1(b))。

図 1 顔領域の抽出

Fig. 1. Extraction of facial area.

この顔領域の抽出は、修正 HSV 表色系⁽⁹⁾によって表される肌色情報と、フレーム間差分に基づく動き情報を基に行っている⁽⁸⁾。

〈2・3〉 頭部の位置の正規化 本方式では、フレーム間差分に基づいて顔部位の変化を捉えているが、撮影された人物の頭部の位置が画像上でずれていないことが前提となる。しかしながら、人物は動くため、カメラと人物の相対的な位置関係が変化し、撮影された人物の画像上の位置や大きさ、向きが変化する。そこで、両目の中心と鼻の頂点の3点の座標で頭部の位置を定義し、これらの座標を基準座標に移動することで頭部の位置を正規化することとした。

頭部の位置の正規化は以下の4段階で構成される。

- (1) 基準座標の設定
- (2) 参照点の抽出
- (3) 参照点の座標補正
- (4) 顔領域のアフィン変換

〈2・3・1〉 基準座標の設定 顔画像をアフィン変換する際に基準とする座標は図2のように設定した。これらの座標は、100人の被験者の顔画像を対象に、まず平常顔の大きさを80,000画素に正規化し、次いで顔領域の重心が画面の中心に来るように平行移動し、その両目の中心、鼻の頂点、および口の両端の座標の平均値として求めたものである。

これらの点は、肌部に比べエッジ成分が多く、部分特徴を用いることで安定に抽出できるとの報告がなされている⁽¹²⁾。

〈2・3・2〉 参照点の抽出 基準点に対応する参照点を、入力された顔画像の各フレームから抽出する⁽¹⁾⁽²⁾⁽⁷⁾⁽¹⁰⁾。ただし、目の参照点としては瞳の中心を抽出するのではなく、目頭と目尻を検出し、それらの中点を参照点として用いることとした。これは、瞳の中心を参照点として用いると⁽³⁾、瞳が正面以外の方向を向いた場合、目の中心からずれてしまうという問題のあることが明らかになったためである。

〈2・3・3〉 参照点の座標の補正 参照点の抽出は各フレーム毎に実施されるため、フレーム毎にその座標が微妙にずれていることがある。そこで各参照点毎に5フレームの座標の移動平均をとることとした。

〈2・3・4〉 顔領域のアフィン変換 ここでは顔部位の変化のみを捉えるために、表情の表出の有無に関わらず、その位置情報が安定している両目の中心と鼻の頂点の3点を用いて頭部の位置の正規化を行う。

この処理手順は次の通りである。

右目の中心	$(x^{e.r}, y^{e.r}) = (302, 212)$
左目の中心	$(x^{e.l}, y^{e.l}) = (414, 212)$
鼻の頂点	$(x_n, y_n) = (358, 274)$
口の右端	$(x^{m.r}, y^{m.r}) = (322, 336)$
口の左端	$(x^{m.l}, y^{m.l}) = (394, 336)$

図 2 基準座標

Fig. 2. Standard points.

- 上記〈2・3・3〉で補正された各フレームから抽出した両目の中心と鼻の頂点の3点を参照点として使用し(図3(a)参照)、これらの座標が基準座標へ移動するようにアフィン変換する(図3(b))。

〈2・4〉 平常顔の検出 平常顔の検出は以下の3段階からなる。すなわち

- (1) 顔部位の変化量の抽出
- (2) 平常顔候補フレームの検出
- (3) 平常顔候補フレームの評価

である。

〈2・4・1〉 顔部位の変化量の抽出 表情の表出に伴う顔部位の変化はフレーム間差分を用いて、顔表面の濃度分布の変化として捉えることで抽出できる。

ただし、頭部が移動した場合は、顔領域全体の濃度分布が同時に変化するため、単に顔領域のフレーム間差分を抽出しただけでは顔表面の変化を抽出するのは困難である。そこで、表情が表出したときに顔部位の変化が大きい目の周辺と口の周辺の2箇所の濃度分布を比較することにより、顔領域全体の濃度分布の変化を除去することとした。これは、表情が表出した場合の目周辺と口周辺の変化には相関性が乏しいが、顔領域全体の濃度分布が変化した場合、目周辺と口周辺の濃度分布の変化に相関がみられると考えられるためである。

この処理手順は次の通りである。

- (1) 目周辺と口周辺に矩形領域を設定する。
- (2) 式(1)と式(2)を用いて、抽出された各矩形領域内で、前フレームと現フレームとの間の平均差分値 M_t^{ae}, M_t^{am} を求める。
- (3) 式(3)を用いて M_t^{ae} と M_t^{am} の差 M_t を求め、変化量とする。

(a) Extracted reference points

(b) Transformed facial area

図 3 頭部の位置の正規化

Fig. 3. Initialization of the head position.

$$M_t^{ae} = \frac{\sum_{x_t^{ae}, y_t^{ae}} |f_t(x_t^{ae}, y_t^{ae}) - f_{t-1}(x_t^{ae}, y_t^{ae})|}{S_t^{ae}} \quad (1)$$

$$M_t^{am} = \frac{\sum_{x_t^{am}, y_t^{am}} |f_t(x_t^{am}, y_t^{am}) - f_{t-1}(x_t^{am}, y_t^{am})|}{S_t^{am}} \quad (2)$$

$$M_t = |M_t^{ae} - M_t^{am}| \quad (3)$$

ここで、フレーム t における目周辺と口周辺の矩形領域内の画素の座標をそれぞれ (x_t^{ae}, y_t^{ae}) , (x_t^{am}, y_t^{am}) とし、矩形領域の面積をそれぞれ S_t^{ae} , S_t^{am} とする。

頭部が動かずに、顔部位のみが変化した場合には原画像の顔領域に対してフレーム間差分を適用すればよいが、頭部が動いた場合には、3点で変換された顔領域を用いる必要がある。

従って、矩形領域の設定は頭部の動きによって以下の2つの処理のいずれかでされる。

(1) 頭部の動きが少ない場合

頭部の位置の正規化処理を行う前の顔領域（図3(a)）に対し、〈2・3・3〉で得られた現フレームの目の中心と鼻の頂点の座標を基に、図4のような、目周辺と口周辺の矩形領域を抽出する。

図 4 顔部位変化を抽出するための矩形領域（頭部の動きが無い場合）

Fig. 4. Rectangular areas for extracting facial movement (without head movement).

$$(x^{ae,r}, y^{e,r}) = (282, 156)$$

$$(x^{ae,l}, y^{e,l}) = (437, 239)$$

$$(x^{am,r}, y^{m,r}) = (304, 300)$$

$$(x^{am,l}, y^{m,l}) = (415, 370)$$

図 5 顔部位変化を抽出するための矩形領域（頭部の動きが有る場合）

Fig. 5. Rectangular areas for extracting facial movement (with head movement).

頭部の動きが少ないため、現フレームから抽出された矩形領域を前フレームにも適用することで、フレーム間での参照点の抽出誤差による影響を避けることが可能である。

ここで、図4の RW と RH は $RW = 20$, $RH = 60$ と設定した。

(2) 頭部の動きが大きい場合

上記〈2・3・4〉の処理で頭の位置が正規化された各フレームの顔画像（図3(b)）に対し、図5に示すような、目周辺と口周辺の矩形領域を抽出する。

正規化された顔領域を用いることで、頭部が移動した場合でも、頭部の動きが除去され、顔部位の変化のみを正確に抽出することが可能である。

〈2・4・2〉 平常顔候補フレームの検出 〈2・4・1〉節で抽出された変化量 (M_t^{ae} , M_t^{am}) の推移を図6(a), (b)に

図6 抽出された変化量の推移

Fig. 6. Transition of extracted movements.

示す。

求められた目周辺と口周辺の変化量 (M_t^{ae} , M_t^{am}) の推移から平常顔候補フレームを検出する。

この処理手順は次の通りである。

- (1) M_t に対し、5点の移動平均を適用し、平滑化を行う。求められた値を M_t' とする (図6(c))。

$$M_{t-2}' = \frac{M_{t-4} + M_{t-3} + M_{t-2} + M_{t-1} + M_t}{5} \quad (4)$$

- (2) 式(5)を用いて、 M_t' の変化が一定時間 (ここでは7フレームとした) 少ない場合を判断し、そのフレーム t を平常顔候補フレームとして検出する。

$$D_{t-n} = |M_{t-n}' - M_{t-n-1}'| < T_{D_t} \quad (5)$$

$$(n = 0, 1, \dots, 6)$$

瞬きを連続的に行う被験者の場合、瞬きと瞬きの間隔は平均10フレームであった。従って、一定時間として10フレームに設定した場合、目の部分に常に変化が捉えられ、平常顔が検出できないという状態が考えられる。そこで、一定時間として7フレームと設定した。

従来⁽⁴⁾は予備実験に基づき $T_{D_t} = 1.0$ と設定し、ある程度の検出精度を得ていた。しかしながら、頭部の動きが極めて小さく、表情表出による顔部位の変化量 M_t' が小さい場合には、表情が表出しているにも関わらず、平常顔候補フレームとして検出してしまいう問題があった。 T_{D_t} の値を小さくすれば、微妙な顔表面の変化を捉えることが可能であるが、頭部に動きがあった場合に、平常顔候補フレームを検出できなくなる可能性がある。そこで、平常顔候補フレームが10フレーム連続して検出された場合には、式(6)を用いて T_{D_t} を自動的に更新することとした。

$$T_{D_t} = \frac{2}{10} \sum_{n=0}^9 D_{t-n} \quad (6)$$

図7 顔部位の差異

Fig. 7. Differences of facial area.

〈2・4・3〉平常顔候補フレームの評価 従来⁽⁴⁾は上記の手法により動画像に含まれる全ての平常顔のフレームを検出しようと試みていたが、表情を認識するためには動画像中から少なくとも1フレームの平常顔が正しく検出されればよい。また、「顔部位の変化が一定時間少ない」という条件で平常顔を検出した場合には、表情が表出した状態で顔部位の変化が少ない場合にも平常顔として検出することがあった。そこで、一度平常顔候補フレームが検出された後、間をおいて再び平常顔候補フレームが検出された場合に、検出された平常顔候補フレームに写っている顔が真に平常顔であるか否かを評価することとした。

この評価は以下のように行う。

- (1) 最初に検出された平常顔候補フレームを平常顔 f_ν であるとする (ただし、平常顔候補フレームが連続して検出された場合はその最後のフレームを f_ν とする)。
- (2) $t > \nu + 1$ で平常顔候補フレーム f_τ が検出された場合、 f_ν と、 f_τ との平均差分値 AD を求める。

$$AD = \frac{1}{S_f} |f_\tau - f_\nu| \quad (7)$$

S_f は顔領域の大きさである。

- (3) $AD \leq T_{AD}$ の場合は f_τ は平常顔であると判断し、新たに f_τ を f_ν とする。
- (4) $AD > T_{AD}$ の場合は f_τ は表情顔であると判断し、引き続き f_ν を平常顔を用いる。

ここで、 T_{AD} を決定するための予備実験を行ったところ、2つのフレームとも平常顔であった場合の平均差分値は約6.5であり、どちらかが表情顔であった場合の平均差分値は約12.0であったが表情が微小な場合に7.3という低い値をとる場合もあった。この結果を基に、表情が微小な場合を考慮し、 $T_{AD} = 5.0$ とした。

〈2・5〉特徴量の抽出 以降の処理では、検出された平常顔 f_ν を基準表情として用いる。

〈2・5・1〉顔部位の正規化 図7(a)から明らかなように、頭の位置を正規化した後でも、人によって口の位置は異なっている。この違いを除去するために、〈2・3・4〉節で変換された顔領域を、変換された平常顔から抽出した口両端の2点を加え、再度変換する。

この処理手順は以下の通りである。

(a) Extracted reference points

(b) Transformed neutral face

図 8 顔領域の正規化

Fig. 8. Normalization of facial area.

図 9 固定マスク

Fig. 9. Fixed masks.

図 10 固定マスク適用例

Fig. 10. Example of application of fixed masks.

- (1) 上記〈2・4・2〉節の処理で平常顔と判定されたフレームの〈2・3・4〉節で正規化された顔画像 f_v から両目の中心と、鼻の頂点、口両端の座標を抽出する (図 8(a))。
- (2) 上記 (1) で抽出された座標が基準座標へ移動するように、平常顔として判定されたフレーム以降のフレームの顔領域 $f_t (t > v)$ をアフィン変換する (図 8(b))。

〈2・5・2〉 特徴量の算出 表情の認識には各顔部位において顔表面の 3 次元的形状がどの程度変化したかを捉えることが必要である。この変化を捉えるために、平常顔と表情顔との間の濃度値の 2 乗誤差を用いることとした。

濃度分布は照明条件による影響を受けると考えられる。そこで、特徴量の抽出前に、平常顔と表情顔の濃度分布を正規化することとした。

特徴量の算出は次の手順で行う。

- (1) 変換された平常顔 $f_v(x, y)$ と表情顔 $f_t(x, y)$ から図 9 の固定マスクを用いて各顔部位 $i(x', y')$ を抽出する (図 10)。
- (2) 式 (8) を用いて、抽出した各顔部位から顔部位の変化量 M_i^t を特徴量として抽出する。

$$M_i^t = \frac{\sqrt{\sum_{x'} \sum_{y'} \{f_v(x', y') - f_t(x', y')\}^2}}{S_i} \quad (8)$$

ここで、 i は顔の各部位を表すインデックスである。

式 (8) において、 S_i は部位 i のマスクの面積を表しており、これで割ることによって特徴量の正規化を図っている。

眉においては、その重心も同時に抽出することで、移動方向と距離も捉えることとした。眉が下方に移動した場合に M_i^t に -1 を掛け、負の値としてこれを表す。

〈2・6〉 表情の認識 〈2・4〉節で検出された平常顔を基準として、〈2・5・2〉節で得られた特徴量を基に表情の認識を行う。認識対象とした表情は、喜び、怒り、悲しみ、驚き、嫌悪、恐怖の 6 基本表情である。

顔には、その人の印象が含まれている。そのため、1 枚の表情顔だけでその表情を認識することは困難であり、平常顔と表情顔と比較し、どの顔部位がどのように変化したかを捉えることが必要であると考えられる。そこで、検出された平常顔を基準として、顔部位の変化を追跡し、表情の推移を認識するシステムを提案する。

提案したシステムは筆者等が静止画を用いて表情を認識するために開発したエキスパートシステム⁽³⁾を動画像に適

図 11 メンバシップ関数

Fig. 11. Membership functions.

表 1 表情の種別

Table 1. Index of expression.

表情	平常	喜び	怒り	悲しみ	驚き	嫌悪	恐怖
e	0	1	2	3	4	5	6

用できるように改良したものである。

主な変更点は以下の通りである。

- (1) 動画像への対応
 - (2) 基本 6 表情および平常顔を認識できるように拡張
 - (3) 認識精度向上のためのルール構成の改良
- このシステムは次の 2 つの認識部から成る。

- (1) 各部位の変化の認識
- (2) 各部位の変化の総合的な判断で実施する総合的認識

〈2・6・1〉顔部位の変化の認識 顔各部位の変化は幾つかのパターンに限定される。例えば、眉では、上に引き上げられる、及び内側に引き寄せられる、内側が引き上げられる等である。各部位の変化がこれらの変化パターンのうち、どの変化に該当するかは判断は予め記述されたルールによって認識する。一方、顔部位の変化は同じ表情でもその程度が異なることがある。このため、特徴量 M_t^i を図 11 のメンバシップ関数で評価することでこの曖昧性を評価することとした。

例えば、ある部位 i の変化量 M_t^i が 0.35 という値であった場合、その特徴量は小 ($L^{i,0}(e)$) という変化の程度との一致度 $\alpha_t^{i,0}(e)$ が 0.25 で、中 ($L^{i,1}(e)$) との一致度 $\alpha_t^{i,1}(e)$ が 0.75 で、大 ($L^{i,2}(e)$) との一致度 $\alpha_t^{i,3}(e)$ が 0.0 と評価される。

ここで、 $e(=0, 1, 2, \dots, 6)$ は表情の種別を表す (表 1)。

尚、各顔部位の変化を的確に判断できるように、メンバシップ関数は各部位ごとに用意している。

〈2・6・2〉顔表情の認識 顔表情の認識はルールに基づいて各部位の変化パターンの組み合わせを総合的に判断して行われる。

例えば、喜びの表情を認識するためのルールの 1 つは図 12 のように記述されている。

このルールは、「喜びの表情では、眉部に大きな変化がなく、下頬と口両端に中または大きな変化が見られる」という特徴を表現したものである。

例えば、左頬 ($i = 21$) の変化が大きい ($j = 2$) ことを判断するための条件 “ $[\alpha_t^{21,2}(1)(M_t^{21}) > 0.5 \mid L^{21,2}(1)]$ ” で

```

IF  $[\alpha_t^{4,0}(1)(M_t^4) < 0.2 \mid L^{4,0}(1)]$  AND
 $[\alpha_t^{4,4}(1)(M_t^4) < 0.5 \mid L^{4,4}(1)]$  AND
 $[\alpha_t^{6,0}(1)(M_t^6) < 0.2 \mid L^{6,0}(1)]$  AND
 $[\alpha_t^{6,4}(1)(M_t^6) < 0.5 \mid L^{6,4}(1)]$  AND
 $\langle [\alpha_t^{19,1}(1)(M_t^{19}) > 0.5 \mid L^{19,1}(1)]$  OR
 $[\alpha_t^{19,2}(1)(M_t^{19}) > 0.1 \mid L^{19,2}(1)] \rangle$  AND
 $\langle [\alpha_t^{21,1}(1)(M_t^{21}) > 0.5 \mid L^{21,1}(1)]$  OR
 $[\alpha_t^{21,2}(1)(M_t^{21}) > 0.1 \mid L^{21,2}(1)] \rangle$  AND
 $\langle [\alpha_t^{24,1}(1)(M_t^{24}) > 0.5 \mid L^{24,1}(1)]$  OR
 $[\alpha_t^{24,2}(1)(M_t^{24}) > 0.1 \mid L^{24,2}(1)] \rangle$  AND
 $\langle [\alpha_t^{25,1}(1)(M_t^{25}) > 0.5 \mid L^{25,1}(1)]$  OR
 $[\alpha_t^{25,2}(1)(M_t^{25}) > 0.1 \mid L^{25,2}(1)] \rangle$ 
THEN Happiness (CF(1, 1) = 1.0)

```

図 12 喜びのルール 1

Fig. 12. Rule 1 for Happiness.

図 13 評価関数

Fig. 13. Evaluation functions.

は図 13 より、入力された一致度 $\alpha_t^{21,2}(1)$ が 0.5 以上であれば適合度 $A_t^{21,2}(1) = 1.0$ となり、 $\alpha_t^{21,2}(1) < 0.5$ であれば、例えば $\alpha_t^{21,2}(1) = 0.25$ であった場合は $A_t^{21,2}(1) = 0.5$ となる。

各条件に対する適合度を評価関数を用いて求めた後、式 (9) を用いて各ルールに対する適合度 $P_t^{e,r}$ を算出し、式 (10) を用いて $P_t^{e,r}$ が最大となったルールに適合したと判断する。

$$P_t^{e,r} = \min_{k,l} \{A_t^{k,l}(e)\} \times \text{CF}(e, r) \quad (9)$$

ここで、CF は各ルールに付加した確信度である。また、 $k(\in i)$ は条件部で用いられた部位のインデックスであり、 $l(\in j)$ は条件で用いられた変化の程度のインデックスである。

各表情毎にルールに対する適合度の最大値 P_t^e を式 (10) で求める。

$$P_t^e = \max_r P_t^{e,r} \quad (10)$$

本方式では、入力された表情に、各表情がどれくらい含まれているかという割合 R_t^e が最大の表情を認識結果として出力する。割合 R_t^e は次式で算出する：

$$R_t^e = \frac{P_t^e}{\sum_e P_t^e} \quad (11)$$

図 14 撮影環境
Fig. 14. Environment of recode.

3. 実験結果および考察

〈3・1〉 撮影条件 撮影はカラー CCD のビデオカメラと 2 台のライトを用いて行った。図 14 に撮影したカメラと照明、被験者の位置関係を示す。これらの距離は約 2m としたが、厳密には定めていない。

喜びの表情を撮影する際は、被験者に以下のように指示をした。

- (1) カメラを見てください。
- (2) 平常顔をしてください。
- (3) 喜んで下さい。
- (4) 平常顔に戻して下さい。

ここで (2) と (3) の指示の間は 3 秒程度の時間をおいた。同様にして怒り、悲しみ、驚き、嫌悪、恐怖の表情も撮影した。ビデオに録画された動画を 720×746 画素、30fps、4:2:2、8bit でワークステーションに取り込んだ。

実験には 10 人の被験者に対して、6 表情を撮影した計 60 本の動画、合計 19,448 フレームを用いた。1 本の動画は短いもので 160 フレーム、長いもので 718 フレームとなった。

図 15 に入力画像の例を示す。

〈3・2〉 平常顔検出実験 この被験者の嫌悪の表情（図 15(e)）を撮影した 180 フレームの動画では、第 55 フレームあたりから表情が表出し、第 120 フレームあたりで平常顔に戻っていた。

変化量の推移の抽出結果を図 16 に示す。平常顔として、第 7-第 16 フレームが検出された。さらに、第 56 フレームが平常顔の評価基準を満たし、平常顔として検出された。

従来の手法⁽⁴⁾では平常顔として、第 8-第 17、第 30、第 44-第 51、第 71-第 91、第 105-第 112、第 125-第 133、第 148-第 150、第 164-第 166 フレームが検出されていた。これは、表情が表出した状態で顔部位の変化がほとんどなかった場合に、その顔を平常顔として誤検出してしまおうという問題があったためである。

図 15 入力画像
Fig. 15. Input video images.

図 16 変化量の推移の抽出結果
Fig. 16. Extraction results of transition of movements.

本手法では、この問題に対応するため、顔部位の変化量の判定基準である閾値 T_{D_t} を自動更新すると共に、検出された平常顔候補フレームの評価を行なっている。

顔領域全体の濃度分布の変化を除去するために、式 (3) では目周辺と口周辺の変化の差をとっているが、これにより表情が表出した場合の目周辺と口周辺の変化までも除去してしまうことが考えられる。実際に、図 16 の第 75 フレームあたりでは、表情が表出している状態で顔部位の変化が少ないために M_t^m が非常に小さな値となった。このため、従来手法では、第 71-第 91 フレームを平常顔として誤検出していた。しかしながら、本手法では、検出された平常顔候補フレームの評価を行ったことで、これらのフレームを誤検出することはなかった。

表 2 に従来手法⁽⁴⁾と本手法の平常顔の検出精度を示す。

表 2 平常顔検出精度
Table 2. Detecting accuracy of neutral face.

	平常顔	平常顔以外
従来手法	3,105 / 6,512 (47.7%)	6,085 / 12,936 (47.0%)
提案手法	1,218 / 6,512 (18.7%)	12,820 / 12,936 (99.1%)

主観評価では用いた 19,448 フレームの内、平常顔が 6,512 フレームで、瞬きや表情顔など平常顔以外のフレームが 12,936 フレームであった。

従来手法では、6,512 フレームの平常顔のうち 3,105 フレームを平常顔として検出し、12,936 フレームの平常顔以外のフレームのうち 6,085 フレームを平常顔以外と判断した。平常顔の検出率は高いが、平常顔以外の検出率が低く、平常顔以外のフレームでもそのフレームを平常顔として検出する場合が多かった。平常顔の検出の目的は、表情認識の際に基準とする基準表情を特定することである。従って、各動画画像から少なくとも 1 フレームが検出されていればよく、各動画画像中に含まれる平常顔のフレームを全て検出する必要は無い。提案手法では、平常顔の検出率は高くはないが、平常顔以外のフレームのほとんどを平常顔以外のフレームであると正しく判断しており、提案手法の有効性が確認された。

また、閾値 $T_{D_i} = 1.0$ で固定していた従来手法では、頭部の移動がほとんどない場合に平常顔が全く検出されない場合が表情顔を記録した動画画像 30 本中 2 本あった。本手法を用いた場合は顔領域の抽出に失敗し、平常顔の検出まで行なわれない例が 2 本の動画画像であったが、それ以外の 58 本については 1 フレーム以上が検出された。しかし、58 本中 1 本の動画画像で微笑んでいる画像が平常顔として検出された。これは微笑んだ状態でのフレームが比較的長く続き、しかもその間での動きが少なかったために、先ず微笑みのフレームが平常顔として検出され、以降のフレームで検出された真の平常顔を平常顔ではないと評価していたことが原因であった。

ここで実験では、予備実験より $T_{AD} = 5.0$ としたが、閾値 T_{AD} の設定について考察する。 T_{AD} の値は既に検出された平常顔と現フレームで検出された平常顔候補フレームの平均差分値の大小を評価するために用いた。この閾値の設定では、現フレームで検出された顔が表情顔ではなく、平常顔であることを判断するため、表情顔が検出されないような値でなければならない。実験で用いた動画画像から無作為に 10 本を選び、平均差分値の分布を調査した (図 17)。比較対象の 2 フレームの顔が平常顔のときの平均差分値を “Neutral” でプロットし、片方の顔が表情顔のときの平均差分値を “Expressive” でプロットした。図 17 をみると、片方が表情顔である場合は 7.0 以上の範囲に分布していることがわかる。従って、 $T_{AD} > 7.0$ に設定した場合、表情顔のフレームが平常顔として誤検出される可能性があることがわかり、今回用いた $T_{AD} = 5.0$ は適当であると考えられる。

図 17 閾値 T_{AD} の評価
Fig. 17. Evaluation of threshold T_{AD} .

図 18 識別結果
Fig. 18. Recognition result.

〈3・3〉表情認識実験 表情の認識は平常顔が検出される毎に行われるが、ここでは平常顔として検出された第 16 フレームと 56 フレームを基準として、表情認識を行なった結果、最大の割合をとった表情を図 18 に示す。この結果、第 56 フレームを基準とした表情の認識では、第 58–第 107 フレームの多くのフレームで嫌悪の割合が最大となった (図 18)。

しかしながら、表情が表出していない第 16 フレームを基準とした第 55 フレームまでの表情認識では全てのフレームを平常と認識しなければならないが、第 21–第 27 フレーム及び第 34–第 47 フレームを驚きあるいは恐怖であると誤認識している。この誤認識の原因に関して詳しく調査したところ、顔領域の正規化で使用する参照点の抽出誤差と正規化方法に問題があることが分かった。

誤認識となった第 21–第 27 フレーム及び第 34–第 47 フレームでは、この被験者は瞬きをしていた。目を閉じた場合に抽出される参照点の位置が下に下がることが確認され、その結果眉の位置が上に上がったかのように正規化された。これにより驚きの表情であると誤認識されていた。

表情識別結果を表 3 にまとめた。これは各動画画像で、識別結果が最大数を占める表情で示している。

4. まとめおよび今後の課題

本論文では、基準顔 (平常顔) の検出アルゴリズムと、表情認識アルゴリズムを提案した。

実験では、顔部位の変化が少ないという特徴を用いて平常顔を検出することができた。また、頭部の動きを抑え、表

表 3 表情識別結果

Table 3. Results of classification.

Input expression	Results of classification					
	Hap.	Ang.	Sad.	Sur.	Dis.	Fea.
Hap.	49/58 (84.5%)	1/58 (1.7%)	2/58 (3.4%)	3/58 (5.2%)	1/58 (1.7%)	2/58 (3.4%)
Ang.	0/60 (0.0%)	47/60 (78.3%)	3/60 (5.0%)	3/60 (5.0%)	5/60 (8.3%)	2/60 (3.3%)
Sad.	0/60 (0.0%)	3/60 (5.0%)	45/60 (75.0%)	5/60 (8.3%)	4/60 (6.7%)	3/60 (5.0%)
Sur.	0/60 (0.0%)	2/60 (3.3%)	3/60 (5.0%)	46/60 (76.7%)	3/60 (5.0%)	6/60 (10.0%)
Dis.	1/60 (1.7%)	7/60 (11.7%)	2/60 (3.3%)	4/60 (6.7%)	43/60 (71.7%)	3/60 (5.0%)
Fea.	0/60 (0.0%)	2/60 (3.3%)	3/60 (5.0%)	7/60 (11.7%)	4/60 (6.7%)	44/60 (73.3%)

情を認識することができた。

表情が表出した状態でも平常顔として検出してしまう場合が従来手法で問題となっていたが、検出された平常顔候補フレームの評価を行うことで大幅に改善することができた。しかしながら、最初に検出された平常顔の評価が含まれていないため、この顔が表情顔であった場合には、極端に精度が下がる可能性がある。そこで最初に検出された平常顔に対する評価基準を検討する必要がある。

また、表情の認識では、顔部位の変化量を特徴量としているが、目の開口度や、皺の方向やその深さを捉えた特徴量を用いることで、さらに高精度な認識が行えるものと考えられる。これについては別の機会に報告する予定である。

また、嫌悪の表情が表れる直前に怒りの表情が表れる等、単に被験者に指示した表情のみが表れるとは限らないことがわかった。さらに多くの被験者の表情の推移を観察することで、様々な表情の特徴を分析し、微妙な表情まで認識できるシステムに生かすことができると考えられる。

謝 辞

本研究に関して様々な議論をしてくださった方々に感謝致します。また、本研究の一部は電気通信普及財団の助成を受けて行った。

(平成 12 年 8 月 8 日受付, 同 13 年 5 月 7 日再受付)

文 献

(1) Hideyuki EBINE & Osamu NAKAMURA: The Recognition of Facial Expressions Based on Production System. Proc. of IEEE CCECE'99. 675-682 (1999)
(2) Hideyuki EBINE & Osamu NAKAMURA: The Recognition of Facial Expressions Based on Production System. Can. J. Elect. and Comp. Eng., Vol. 24. 99-107 (1999)
(3) 海老根 秀之, 中村 納: 「顔部位の個人的差異を排除した顔表情の認識」, 電学論 C, Vol. 119-C, No. 4, 474-481 (1999)
(4) 海老根 秀之, 志賀 泰博, 池田 美鈴, 中村 納: 「表情認識における基準表情の自動検出と表情推移の認識」, 電学 C 大, OS5-2, 355-360 (2000)
(5) P. Ekman & W. V. Friesen: UNMASKING THE FACE. Prentice Hall, Inc., (1975)
(6) Ifran A. Essa & Alex P. Pentland: Facial Expression Recognition using a Dynamic Model and Motion Energy. Proc. of

IEEE Int. Conf. Compt. Vis., Vol. 5. 360-367 (1995)
(7) 福井 和広, 山口 修: 「形状抽出とパターン照合の組合わせによる顔特徴点抽出」, 信学論 D-II, Vol. J80-D-II, No. 8, 2170-2177 (1997)
(8) 池田 美鈴, 海老根 秀之, 志賀 泰博, 中村 納: 「動画像からの顔領域抽出法」, 信学全大, Vol. D-12-58, 228 (2000)
(9) Satoshi MATSUHASHI & Osamu NAKAMURA & Toshi MINAMI: Human-face extraction using modified HSV color system and personal identification through facial image based on isodensity maps, Proc. of IEEE CCECE'95, Vol. 2, 909-912 (1995)
(10) 大津 展之: 「判別および最小 2 乗基準に基づく自動しきい値選定法」, 信学論 D, Vol. J63-D, No. 4, 349-356 (1980)
(11) 大塚 尚宏 & 大谷 淳: 「時系列表情画像の解析処理に基づく表情変換法の検討」, 信学技報, Vol. PRMU96-202, 75-80 (1997)
(12) 宋 欣光 & 李 七雨 & 徐 剛 & 辻 三朗: 「部分特徴テンプレートとグローバル制約による顔器官特徴の抽出」, 信学論 D-II, Vol. J77-D-II, No. 8, 1601-1609 (1994)

海老根 秀 之 (学生員) 1997 年 3 月 工学院大学工学部電子工学科卒業。1999 年 3 月 同大学院情報工学専攻博士前期課程修了。現在, 同大学院情報学専攻博士後期課程在学中。画像処理, パターン認識の研究に従事。電子情報通信学会, 電気学会各学生会員。

志 賀 泰 博 (非会員) 1999 年 3 月 工学院大学工学部電子工学科卒業。2001 年 3 月 同大学院情報学専攻修士課程修了。現在, 凸版印刷株式会社 金融・証券事業本部勤務。

池 田 美 鈴 (非会員) 2000 年 3 月 工学院大学工学部情報工学科卒業。現在, 同大学院情報学専攻修士課程在学中。画像処理, パターン認識の研究に従事。電子情報通信学会学生会員。

中 村 納 (正員) 1969 年 3 月 工学院大学工学部電子工学科卒業。1971 年 3 月 同大学院修士課程修了。同年, 同大学助手となる。以来, 学習理論, 画像処理, 画像符号化等の研究に従事。1988 年 4 月~1989 年 3 月, ケベック大招聘教授。1992 年工学院大学教授。工学博士。画像電子学会, 電子情報通信学会, 情報処理学会, 人工知能学会, 映像情報メディア学会, IEEE. Pattern Recognition Society, 日本顔学会, 電気学会各会員。