顔の表情の自動認識*

中 村 納**

Automatic Recognition of Facial Expression

Osamu NAKAMURA

1. はじめに

人間同士のコミュニケーションにおいて、意思、感情などは言葉、表情、身振り、手振りなどによって伝達されている。言葉を除いては、そのほとんどは視覚情報によるものであり、視覚情報は多くの民族、人種および文化を越えた共通の認識手段である。その中でも、表情による情報伝達は、人にとって無意識に行われる最も自然なコミュニケーションの形態であり、感情を表出する非常に重要な役割を果たしている。

このような観点から、コンピュータに顔表情を認識させることは人と機械の意思疎通を円滑に進めるマンマシンインタフェースを構築する上で重要であるといえる。これを実現することにより、TV電話、TV会議システムなどにおける顔画像の知的符号化、コンピュータグラフィックスにおけるアニメーションの表情表出などへの応用が期待できる。さらに、この表情の理解によって相手の意思、感情、気分などを正確に知ることができれば、人間同士の非言語的で高度なコミュニケーションを確立することができる。また、単に工学の分野だけではなく、医学、民族学、心理学など他の学問分野に対しても応用が可能となる。

表情に関する研究は、医学、生物学、精神医学、心理学、工学などさまざまな分野で盛んに行われている^{1)~3)}。EkmanとFriesen^{4),5)}は、表情と感情を対応付ける研究を行い、幸福、悲しみ、怒り、恐怖、嫌悪、驚きの6つの基本感情を見出し、それ

ぞれの感情について顔の変化がどの部分にどのように表れるかを整理し、さらに6つの感情が混ざりあった33種類の感情がどのように顔に表れるかについても言及している。そして、表情の変化をリスト構造として記述するFACS (Facial Action Coding System) を構築している。

これに対し、工学の分野においても情報化社会の発展に伴って、工学的な応用を第一の目的として表情の認識、解析に関する研究が行われつつある^{6)~12)}。これらの表情認識に関する多くの研究では、顔の部分構造、すなわち、眉、目、鼻、口などの輪郭から抽出した特徴点を用い、これら特徴点の2次元的な配置関係および動き情報を基にして、顔それ自体、または顔表情を記述するという方法であった。しかし、顔の表情は単にこれら特徴量の物理的な動きだけで決まるものでない。例えば、能面はそのわずかな動きで表面の陰影を変え、この陰影の変化によって表情を変えることができる。一方、歌舞伎などでは陰影表情情報よりもむしろ、色情報を積極的に用いることで表情を出している。すなわち、表情には単に特徴点の物理的動きだけでなく、顔表面の濃度分布、色分布が大きく関わっているということになる。したがって、表情を解析するためのモデルには、単なる特徴点の物理的な動きだけでなく、顔表面の濃度分布の変化や色分布の変化をも記述できるものでなければならない。

筆者らは、工学の立場からアプローチすることであることを明確にして、上記問題点に対処することを念頭としている。すなわち本研究では、コンピュータによる人物の顔表情の自動認識を最終的な

*原稿受付 1995年10月12日

**工学院大学電子工学科

目標に、顔画像から表情を表す特徴量を自動的にかつ安定に抽出できる手法の確立およびこの特徴量を用いて起こり得るすべての表情を記述することのできる表情記述法の確立目的としている。

本方式は、次のような基本的考えを背景として検討されたものである。

- (1) 表情を記述する際の記述パラメータは、単に顔の部分構造の輪郭部から抽出された特徴点だけでは不十分であり、顔全体から抽出されなければならない。
- (2) 記述モデルの中に顔の3次元構造が反映されなければならない。
- (3) 表情の変化がモデルの変化として十分表現されなければならない。

以上の条件を前提として、本論文では、表情を表す特徴量に“差分画像の等濃線分布”を用いた表情記述および認識方式を提案する。等濃線^{13),14)}は地形図の等高線、天気図の等温線、等圧線に対応するもので、画像中の濃度境界線を結んだ閉曲線である。等濃線分布は、その中に顔の3次元的な形状変化が良好に反映されており、また、その抽出が容易である、等の利点を備えている。

以下、2章で本方式の概要について述べ、3章で表情の記述に用いる特徴量の抽出方法について説明する。また、4章で記述された表情を表情空間に表し、5章で本方式の今後の課題について整理する。

2. 本方式の概要

本方式では、表情変化に伴う濃度分布の変化を直接的に捉えるために“表情変化前（平常顔）”の画像と“表情変化後（表情顔）”の画像の間の差分画像に対し等濃線分布を作成し、その等濃線分布から得られる等濃線の密度情報を要素として持つ特徴ベクトルによって表情を記述する。

本方式の概要を以下に示す。また、図1に本方式の流れを示す。

(1) 前処理

入力された画像に対して前処理で、顔領域を抽出し、大きさと位置の正規化を行う。

(2) 差分画像の作成

前処理で得られる表情変化前後の画像それぞれ

図1 本方式の流れ

から差分画像を作成する。このとき差分画像は表情変化前の画像から表情変化後の画像を差し引くため、各画素の濃度値は正あるいは負の符号を持つことになる。差分画像の濃度値が正負どちらの値をとるかは各表情、あるいは、顔の各部位によって異なるため本方式では、これら正・負の符号を別々に扱い、2枚の差分画像を作成することとする。

(3) 等濃線分布の作成

等濃線分布は画像に対して濃度量子化を行った後、量子化画像の濃度境界線を追跡して作成される。まず、正・負それぞれの差分画像に対して濃度ヒストグラムの面積を8等分するようなしきい値を用いて8レベルに量子化する。次に、量子化画像の濃度境界線を追跡して正・負それぞれの差分画像に対する等濃線分布が作成される。このとき、差分値が最も小さい量子化レベルの境界線を第1レベルの等濃線と定義し、差分値が最も大きい量子化レベルの境界線を第8レベルの等濃線と定義する。

(4) 特徴ベクトルの作成

正・負それぞれの差分画像に対して作成された等濃線分布を対象に、別々に密度情報を抽出して特徴ベクトルを作成する。まず、顔の各部位を指

定する部位マスクを設定し、各部位マスクを 8×8 画素あるいは 4×4 画素のブロックに分割する。ついで、各ブロックごとにその中に含まれる最大レベルの等濃線から最小レベルの等濃線までのレベル数を密度情報として抽出し、この各ブロックの密度値を要素とするベクトルを特徴ベクトルとする。ブロックの大きさは額、鼻、両頬、頤の部位では 8×8 画素とするが、表情を認識する上で特に重要と考えられる両目（眉の部分も含む）と口の部位においてはより詳細な情報を得る目的からブロックサイズを小さくし 4×4 画素とする。差分画像の正・負の別、顔各部位別に抽出された各ブロックの密度値を要素とするベクトルを、表情を表す特徴ベクトルとする。

(5) 特徴ベクトルのノルムによる表情の記述

差分画像の正・負の別および顔各部位別に抽出された特徴ベクトルのノルムをそれぞれ求めて表情のモデルを作成する。このとき、ある表情に対し、その表情変化に伴って変化の現れる顔部位および正または負の特徴ベクトルを選んで使用することとする。

3. 特徴量の抽出

本論文で使用する画像は、被験者の正面から一様な光を当てて撮影した正面顔画像で、横 $256 \times$ 縦 240 画素、 256 階調の白黒濃淡画像として入力される。被験者には、カメラに正面を向いて表情を変えてもらい、平常顔と表情顔をそれぞれ入力する。表情としては、FACSに示されている6つの基本感情（幸福、悲しみ、怒り、恐怖、嫌悪、驚き）の中で、表情を表出させる際に比較的イメージしやすい喜び（幸福）、怒り、驚き、悲しみ、の4表情だけを用いることとした。また、背景に関して

は、本研究の目的とは直接関係なく、また前処理の負担を軽減するという目的から今回の実験では暗幕を用いることとした。入力画像の例を図2に示す。

以下、特徴量の抽出方法について述べる。

3.1 前処理

前処理部では、入力画像から背景、髪、頤から下の領域を含まない顔領域の抽出を行う。また、顔領域の大きさについてはその面積が $20,000$ 画素（画像全体は $61,440$ 画素）に正規化する。図3に顔領域を切り出した画像（(a)平常顔、(b)表情顔）の例を示す。さらに顔領域の位置の補正を行う。顔画像の入力時には、できるだけ頭の位置を動かさないよう指示しているが、完全に固定することは不可能である。このような微妙な顔の移動は、差分画像において表情変化に無関係な濃度分布の変化となって現れる。したがって、これらの誤差要因を除くために位置補正が必要となる。そこで、表情変化等においてもそれほど形状に変化のない鼻領域をテンプレートとして用い、テンプレートマッチングによって、位置の正規化を行うこととした。平常顔に対する鼻領域のテンプレートマッチングでは、すべての被験者に対して共通のテンプレートを用い、マッチした鼻領域が画面の中央にくるように位置の正規化を行う。ただし、各被

図3 顔領域の抽出例

図2 入力画像の例

図4 差分画像（絶対値）の例

験者で鼻の形状や濃度分布が異なることを考慮し、表情顔に対する鼻領域のテンプレートマッチングでは、各被験者の平常顔の鼻領域をテンプレートとして用いている。図4に位置補正を行わなかった場合の差分画像の例を示す。位置補正なしでは目や眉の部分で位置ずれの影響が認められる。

3.2 差分画像の作成

本方式では表情の変化を直接的に捉えるために平常顔と表情顔の差分画像を作成する。差分画像は平常顔の濃度値から表情顔の濃度値を差し引くため、平常顔において濃度値が高く（暗い領域）、表情顔において低い（明るい領域）画素は、濃度値の差分の符号は正となる。逆に、平常顔において濃度値が低く（明るい領域）、表情顔において高い（暗い領域）場合、差分値は負の符号になる。表情変化において、この正の部分と負の部分は顔部位の別々の変化を表している。例えば、歯の出現は正の部分で特徴が現れ、“しわ”の出現は負の部分で特徴が現れる（表1に表情変化による顔部位の変化とその変化が現れる正、負の別を示す）。したがって、本方式ではこの正の部分と負の部分

図5 差分画像の例

をそれぞれ分けて、2枚の差分画像を作成する。図5に差分画像（(a)正の部分、(b)負の部分）の例を示す。

3.3 等濃線分布の作成

等濃線分布は、画像を量子化し、その濃度境界線を追跡することによって得られる。本方式では前記3.2で得られた2枚の差分画像に対して、等濃線分布をそれぞれ作成する。

差分画像の濃度情報を大まかに捉えるため、画像を8レベルに濃度量子化する。しかし、照明環境などの撮像条件を、一定に保つことは困難であり、また、被験者によって顔の濃度値の高低差は不定である。すなわち、そのときの照明条件や被験者の顔濃度の個人差などの影響を考慮すると、量子化のしきい値を固定することはできない。そこで本方式では、差分画像の濃度ヒストグラムの面積を8等分するようなしきい値を用いて各画素の濃度値を8レベルに量子化することとした。すなわち、画像中における各量子化レベルの画素数はほぼ等しくなる（図6）。

3.4 特徴ベクトルの作成

作成された差分画像の等濃線分布を、 8×8 画素あるいは 4×4 画素のブロックに分割し、各ブロックごとにその中に含まれる最大レベルの等濃

表1 顔部位の変化とその表情が現れる差分画像の正・負の別

表情	顔部位	+	-	顔部位の変化
喜び	眉	±		変形する
	目	±		変形する
	目尻		-	しわが出現する
	鼻、頬	-		鼻から口の両端に走るしわが出現する
	頬上部	+		頬が持ち上がる
	頬下部	-		頬がこける
	口上部	-		上唇が移動してくる
	口中央	+		歯が出現する
		-		口が開く
	口下部	-		下唇が移動してくる
	口の端	±		移動、変形する
怒り	眉	±		移動する
	目	±		変形する
	眉間	-		しわが出現する
	口の端	±		変形する
驚き	額	-		しわが出現する
	眉	±		移動する
	目	±		変形する
	頬上部	+		頬が持ち上がる
	頬下部	-		頬がこける
	口上部	-		上唇が移動してくる
	口中央	-		口が開く
	口下部	-		下唇が移動してくる
	口の端	±		移動、変形する
悲しみ	額	-		しわが出現する
	眉	±		移動、変形する
	目	±		変形する
	眉間	-		しわが出現する
	口の端	±		変形する

図6 等濃線分布の例

線から最小レベルの等濃線までのレベル数を密度情報として抽出する。この密度情報には、濃度差分の変化の度合いが示されていると同時に、3次元形状の変化の度合いも反映されている。各ブロックの密度情報を要素とするベクトルが特徴ベクトルとなる。

ここで考慮しなければならないのは、各被験者ごとに顔における眉、目、鼻、頬、口など、顔部位の相対的な位置関係が異なるということである。このような顔部位の位置情報は個人の特徴を表しているが表情の表出には関係ない。したがって、部位マスクを設定する際には各被験者ごとに適切な位置に部位マスクを設定する必要がある。そのため、部位マスクは大まかに額、右目、左目（眉の部分も含む）、鼻、右頬、左頬、口、頤の8つの顔部位ごとに別々に設定することとし、顔部位の位置決定はテンプレートマッチングで行うこととした。ただし、額と頤の部位については、テンプレートが作りにくいという理由の他に、顔領域についてはその大きさと位置の正規化を前処理段階ですで行っているため、位置に関してそれほど個人差はないと判断できることからすべての被験者に対して位置を固定することとした。また、顔領域の位置補正において鼻領域を基準にしていることから、鼻の部位についても位置を固定している。図7に顔の各部位の形状と位置を示す（ただし、目、頬、口はテンプレートマッチングで位置を決定しているため、実際は各被験者ごとに図より若干ずれることがある）。テンプレートマッチングでは太い枠で囲まれた領域ごとに部位が抽出され、さらに、表1に基づいて、眉を含んだ目の領域は眉、目、目尻に、鼻の領域は眉間、鼻に、頬

図7 顔部位の設定

の領域は頬上部、頬下部に、口の領域は口上部、口中央、口下部、口の端に細かく分割される。この時、それぞれの領域において位置は固定とした。

テンプレートマッチングは平常顔の顔領域画像（図3(a)）に対して行い、画像におけるその顔部位の位置情報を等濃線分布画像に当てはめて用いる。その後、顔各部位ごとにブロック分割を行う。顔各部位の中で表情を認識するのに重要と考えられる目（眉の部分も含む）と口の部位については細かく情報を得るために4×4画素にブロック分割し、その他の部位、すなわち額、鼻、頬、頤については8×8画素のブロックに分割した。図8に密度情報を画像に表した例を示す（(a)差分が正の部分、(b)差分が負の部分）。図でブロックの濃度がそのブロックの密度を表している。なお、顔部位の位置の決定は各部位ごとに独立に実施しているため、その領域は重なる部分が生じる場合もある。

以上で、正・負別々の差分画像の等濃線分布から顔各部位ごとに合計38個の特徴ベクトルが作成される。

3.5 特徴ベクトルのノルムによる表情の記述

差分の正・負別々に抽出された38個の特徴ベクトルについて、表1に示した調査結果に従って顔部位の変化に特に関係の深い特徴ベクトルのノルムを求め、その相対的關係で表情を記述する。なお、表1で正と負の両方に関係があると判断された部位については正、負のベクトルを合わせて求めたノルムを用いている。したがって、顔表情は19個のパラメータによって記述されることになる。図9に表情の記述例を示す。

4. 表情空間

前章で記述された表情を視覚的に表すため、本章では表情空間を作成する。19個の特徴ベクトルのノルムによって表された表情は、19次元空間上

図9 表情の記述例 (入力画像は図2)

における一点として表せる。したがって、記述された表情は19次元の表情空間の一点で表されることになる。図10に表情空間を幾つかの入力画像の例とともに平面に表した図を示す。19本の軸は特徴ベクトルのノルムであるが、表1に表したように同時に顔各部位の変化の種類でもある。

4.1 標準パターンについて

表情空間においては、特徴ベクトルのノルム、すなわち各部位の変化が軸となっている。そのため別々の入力画像において顔各部位の変化の差が小さければ、表情空間中でも互いに近い位置に存在することになる。すなわち、似ている表情は表情空間中でも互いに近い位置にままと考えられる。そこで、図10には比較のために、今回使用した4表情 (喜び、怒り、驚き、悲しみ) の標準パターンを示した。例えば、同じ喜びの表情の中でもさまざまなものが存在するが、その中から代表的なものを幾つか選び、喜びの標準パターンを作成した。各表情ごとに、一つの特徴ベクトルに注目し、そのノルムについてヒストグラムを作成すると、幾つかのサンプルは図11のようにある範

図10 表情空間

図11 各表情のノルムのヒストグラム(眉間・負)

囲に集中する傾向がある(この場合は眉間の部位で負の差分画像から抽出した特徴ベクトルである)。このようにある範囲に集中しているものを代表的な喜びのパターンとして選び、それぞれの特徴ベクトルのノルムの平均を求めて標準パターンとした。図12から図15に標準パターンとして用いた表情の例を、図16から図19に作成された標準パターンを示す。また、標準パターン同士の距離を表2に示す。表2から怒りと悲しみが比較的近い位置にあるのが分かる。

4.2 各表情の分布について

表情空間において、喜びの表情では口に大きな変化があり、驚きの表情では額や眉に大きな変化があるため原点より離れた分布になり、また、怒りと悲しみの表情では比較的变化が小さいために、それぞれ原点寄りに分布している。また、本方式

図12 喜びの標準パターンを作成するために用いた表情の例

図13 怒りの標準パターンを作成するために用いた表情の例

図14 驚きの標準パターンを作成するために用いた表情の例

図15 悲しみの標準パターンを作成するために用いた表情の例

図16 喜びの標準パターン

の記述では各表情の分布の広がりとは同じではなく、むしろ異なっていると考えられる。さらに、各表情の境界がどこになるかは見る人間あるいは表情

図17 怒りの標準パターン

図19 悲しみの標準パターン

図18 驚きの標準パターン

表2 標準パターン同士の距離

	喜び	怒り	驚き	悲しみ
喜び	0.00	14.50	21.71	22.14
怒り	14.50	0.00	14.88	12.18
驚き	21.71	14.88	0.00	15.12
悲しみ	22.14	12.18	15.12	0.00

表3 各表情の標準パターンからの分布範囲

	喜び	怒り	驚き	悲しみ
分布範囲	22.34	19.18	19.24	18.19

を表している本人によって異なるとも考えられる。

そこで、それぞれの表情がどのような範囲に分布しているのかを調べるため、各表情のサンプルが含まれる標準パターンからの距離を調査した。ただし、サンプルがどの表情に属するのかは筆者らの主観による。表3のように喜びでは約22、怒りでは約19、驚きでは約19、悲しみでは約18の標準パターンからの距離に各サンプルが含まれていた。

ここで、各表情におけるサンプルの標準パターンからの分布が異なる表情の標準パターン同士の

距離よりも大きくなっているが、これは、標準パターンの作成時に選出したサンプルに若干問題があったことと、実際に使用した表情が極めて変化に富んでいて、各表情の境界付近に存在し、主観的判断でも正確に分類することが困難な表情が含まれていたことに起因していると考えている。

また、このように分布が広がった原因としては、個人の特徴が含まれたためだと考えられる。すなわち、標準パターンはいくつかのサンプルを平均して作成したために個人の特徴が薄れ、各表情で近い位置に分布しているが、個々のサンプルでは個人の表情の表し方の特徴が存在しており、標準パターンから遠い距離にまで分布したと考えられる。喜びの分布が他の表情に比べて広いのは、表情の表し方の個人差が他の表情よりも大きいためとも考えられる。

5. 今後の課題

今後の検討課題を以下に整理する。

(1) 表情の数値化

任意の表情が喜び、怒り、驚き、悲しみの標準パターンとどの程度一致しているかを数量的に表し、表情認識に応用する。この時、認識結果はできるだけ人間の主観と近い結果になるようにする。

例えば、任意の表情を、喜び、驚きとの一致度がそれぞれ0.4ずつで、怒り、悲しみのそれぞれとの一致度が0.1ずつであるというように表す。この場合で、喜びと驚きは0.4ずつで等しいが、多くの人が喜びの表情だが驚いているようにも見えたと判断すれば、同じように（例えば20%）驚きの表情の含まれた喜びの表情と認識結果を出力できるようにする。

(2) 特徴量の抽出法の検討

本方式では、特徴ベクトルとして差分画像の等濃線分布から密度報を抽出して用いている。しかし、この等濃線分布を作成する際のしきい値、レベル数についての詳細な調査は現在検討段階である。今後は、重要となる濃度分布の変化としきい値、レベル数などの関係を調査する予定である。

(3) 入力時の環境条件の影響の調査

本方式では、表情を表す特徴を平常顔と表情顔の差分画像から抽出している。このため、顔画像を撮影する際における照明条件、または、顔の傾きなど入力時の撮像条件による顔の陰影の変化が直接的に密度情報に現れていると考えられる。そこで、純粋な表情の変化と環境条件による変化との違いを調査し、撮像条件の影響を取り除く手法の検討が必要である。

6. まとめ

本論文では、コンピュータによる顔表情の自動認識を目標として、差分画像の等濃線分布に基づく表情の記述について述べた。

表情変化に伴う濃度分布の変化を直接的に捉えるために表情変化前（平常顔）と表情変化後（表情顔）の差分画像から等濃線分布を作成し、単位

領域（1ブロックが 8×8 画素または 4×4 画素）内における等濃線レベルの密度情報を抽出して特徴ベクトルを作成した。さらに、表情を表す特徴量としてその特徴ベクトルのノルムを利用して表情を記述し、4表情（喜び、怒り、驚き、悲しみ）を表情空間に表した。

記述された表情を人間の感性と合った認識、例えば多くの人が喜びに近い表情と感じる表情ならば、同じように喜びに近い表情と認識し、この表情に意味のある解釈を与えるのは今後の課題となる。

最後に、本稿をまとめるに当たりご援助いただいた工学院大学南敏教授に厚く御礼申しあげる。

参考文献

- 1) 桐田：表情の認識—心理学の立場から—, MEDICAL IMAGING TECHNOLOGY, Vol.12, No.6, pp.681-687 (1994-11).
- 2) 森島：表情の認識—工学の立場から—, MEDICAL IMAGING TECHNOLOGY, Vol.12, No.6, pp.688-693 (1994-11).
- 3) 角辻：精神科医からみた顔画像認識への期待, MEDICAL IMAGING TECHNOLOGY, Vol.12, No.6, pp.700-709 (1994-11).
- 4) Paul Ekman, Wallace V.Friesen: Unmasking the Face, Prentice-Hall, Inc., Englewood Cliffs, (1975). (工藤力訳, 表情分析入門, 誠信書房, 1987).
- 5) Paul Ekman, Wallace V.Friesen: Facial Action Coding System, Consulting Psychologists Press, Inc., Palo Alto, (1978).
- 6) 崔, 原島, 武部：顔の3次元モデルを用いた顔面表情の分析, 信学論(D-II), Vol.J74-D-II, No.6 pp.766-777 (1991).
- 7) 森島：ヒューマンインタフェースのための感情空間表現と表情記述, 重点領域研究「感性情報処理」D班平成5年度第3回研究会, 1, pp.1-6.
- 8) 原, 小林：顔ロボットによる6基本表情の動的表出とその特性, 重点領域研究「感性情報処理」D班平成6年度第2回研究会, 4, pp.21-27.
- 9) 間瀬, 前田, 末永：表情動画像からの感情の認識の一手法, 信学技法, PRU91-24 (1991).
- 10) 松野, 李, 辻：ポテンシャルネットとKL展開を用いた顔表情の認識, 信学論(D-II), Vol.J77-D-II, No.8 pp.1591-1600 (1994).
- 11) 永田, 金子, 原島：顔の印象学に向けて—コンピュータによる解析—, 重点領域研究「感性情報処理」D班平成6年度第2回研究会, 2, pp.9-13.
- 12) Yaser Yacoob, Larry Davis: Computing Spatio-Temporal Representations of Human Faces, IEEE CVPR, Vol.1994, pp.70-75 (1994).
- 13) O.Nakamura, S.Mathur, T.Minami: Identification of Human Faces Based on Isodensity Maps, Pattern Recognition, 24,3, pp.263-272 (1991).
- 14) 上野, 加藤, 中村, 南：等濃線分布に基づく正面顔画像の識別, 信学論(D-II) Vol.J76-D-II, No.3 pp.494-506 (1993).