

## 2-3 表情の自動認識

間瀬 健二\*

### 1. ま え が き

人間対人間のコミュニケーションにおける情報伝達メディアには、言葉のほかに表情、視線、ジェスチャ、姿勢、距離、外観など多くの種類がある。とりわけ、人間対人間の会話において、顔の表情は言葉によるコミュニケーションチャンネルを補ったり、それだけで相手にメッセージを送ることができる。表情によるコミュニケーションは、シンボル化された表情の意図的な生成によることもあるが、感情や情動の自発的な表出であったり、表情を作ることで感情を生起・増幅することもあるため、言葉に現れない心的状態をモニタすることが可能と考えられる<sup>1)</sup>。

表情の科学的分析は Duchenne や Darwin 以来、解剖学、心理学、動作学など多くの分野で行われてきた<sup>2)</sup>。そして、上記のようなコミュニケーションメディアとしての研究が進むなかでコンピュータ技術と結合して、人間対人間だけでなく、人間対コンピュータあるいはコンピュータが介在した人間対人間のコミュニケーションにおけるメディアとして最近工学的にも注目されるようになった。

本稿では画像処理による表情認識技術の歴史と最近の動向についてまとめる。以下、2節では表情認識研究の推移を、3節では認識技術の概要を解説し、4節では最近の技術動向として3つの研究報告を取り上げ紹介する。

### 2. 表情認識研究の推移

表情を分析する研究は、とくに神経解剖学や心理学の分野で長い歴史がある。近年、Ekman と Friesen は感情の表出に関連した顔面の動きを測定して記述する目的で、表情をコードとして表現する FACS (Facial Action Coding System)<sup>3)</sup>を作った。これは顔面筋の位置を考慮していることと、長年の心理学的

な実験や調査に基づいていること、さらに、1つのまとまった体系をもっていることから、コンピュータグラフィックス (CG) による表情生成で使われた<sup>4)</sup>のを契機に、表情のひとつの記述法として多く用いられた。

また、1980年代後半頃からCGのハード/ソフトの性能が向上して、CGによる表情画像の高速生成が可能となった。そこで分析と合成を組合せたモデルベース符号化合成のための表情画像の自動分析の研究が多く行われた (例えば5)。現在、この研究の流れの一部は表情のモーションキャプチャ<sup>6)</sup>とCGによる再現という形に移行し、CGプロダクションなどで多く利用されている。また、3次元CG技術を駆使した臨場感通信会議のようなサイバースペースに対し、そこへ送り込む自分の分身であるアバター (avatar) への変身<sup>7)</sup>手段を与えることによって人間対人間のコミュニケーションにコンピュータが介在した、新しいコミュニケーション形態の提案に一役買っている。

一方、人間対コンピュータのコミュニケーションのためには、表情を感情とか意図といったアブストラクたな表現として認識理解させる必要がある。1990年代初頭には、画像処理で獲得した顔の変形特徴をFACSのように体系づけられた表情要素記述に変換したり、さらに上位の記述である基本表情パターンのカatalogとのマッチングをして、基本表情の認識が可能であることがいくつか報告された。

米国科学財団 (NSF, National Science Foundation) は顔画像認識を含む表情理解研究に興味を示し、1992年7月には顔および表情の理解に関連する心理学者とコンピュータ科学者を集め、表情理解に関するワークショップを開催し、現状報告と今後の提言をまとめさせている<sup>8)</sup>。ここでは心理学とコンピュータ科学の協力研究体制の確立、照明/姿勢/形状の変化への対応という研究課題、標準データベースの整備などが提言された。その後、個人識別、表情認識、ジェスチャ認識などの人物画像処理の研究が世界的に加速し、1995、1996年には顔とジェスチャの自動認識に関する国際会議が開かれるまでになっている (第3

\* 株式会社 ATR 知能映像通信研究所  
"Review on Automatic Facial Expression Recognition" by Kenji Mase (ATR Media Integration & Communications Research Laboratories, Kyoto)

回は1998年に奈良で開催される)。

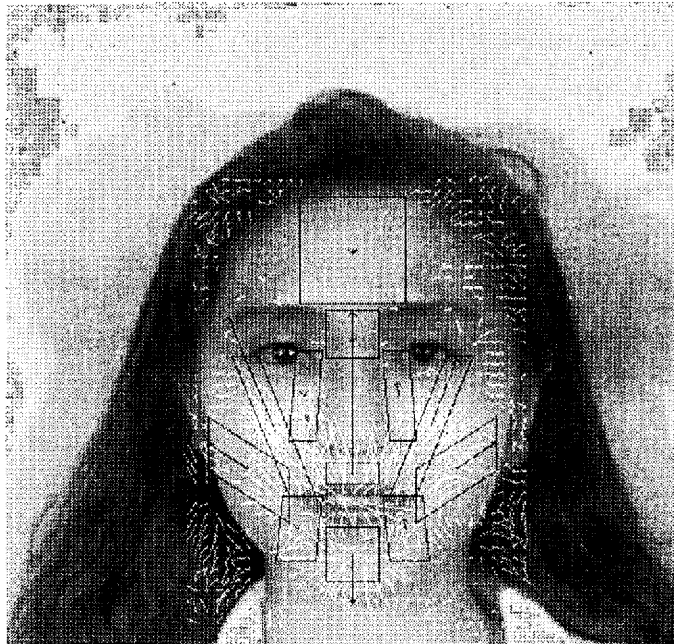
### 3. 表情認識技術の概要

表情認識という課題はもっぱら、その個人にとって無感情の顔からの変形をもとに判断することを議論している。1枚の無感情の写真にもいろいろな表情を読み取ることはできるが、ある基準に対する変形を狭義の表情として捉える。その意味では感情(情動)による表情の表出を認識している、ということもできよ

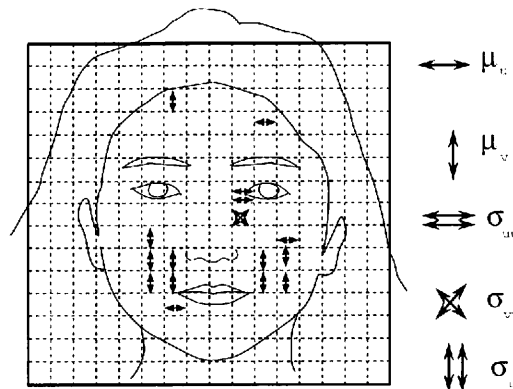
う。すると、表情画像から感情を認識する過程は、次の5段階に分けることができる。

- (1) 顔の造作の抽出
- (2) 顔の変形の検出
- (3) 変形の表情要素による記述
- (4) 基本表情の認識
- (5) 複雑な感情(意図)の認識

造作の抽出は、ほとんどの場合、顔画像識別の手法を流用することができる。表情認識手法のポイント



(a) オプティカルフローから表情筋の動きを推定している



(b) 表情の主たる変形特徴が現れる領域の例

図1 表情認識の2つのアプローチ:(a)表情記述モデルの適用と、(b)時空間画像パターンの特徴ベクトル化

は、(i)造作を含む顔の変形特徴として何を用いるか(変形特徴)、(ii)その変形をどのように記述するか(表情記述)、そして、(iii)記述された変形を表情や感情へマッピングするためにどの手法を用いるか(認識モデル)、ということになる。

これらの組合せは多種多彩で、いろいろなアプローチが考えられる。ここではまず、2つの基本的な例を紹介する<sup>9)</sup>。

第一は、FACSのような表情を記述するモデルで変形を記述する方法(認識モデル)である。表情筋の動作をオプティカルフロー(変形特徴)により推定し、それをFACSのAU(Action Unit)で記述(表情記述)している。主な筋肉の場所に検出窓を置き、その窓内のフローのデータから筋肉の動きを推定する。実際の表情画像について実験を行い、表情を作るための筋肉の動きを検出できることが確かめられている(図1(a)参照)。ここでは勾配法によるオプティカルフローが頬など造作のない領域でも変形を検出できることが注目される。

第二は、表情画像から得られる変形パターンに一般的なパターンマッチングの手法を適用して感情を識別する方法である。前者と同様に、変形特徴として勾配法で得られるオプティカルフロー成分の時空間統計量を用いる。学習サンプルの変形パターンから特徴ベクトルを決定すると、表情画像から有意な場所(図1(b)参照)と特徴量を抽出できる(表情記述)。この学習サンプルを辞書としてパターンマッチングによる識別器にかけると、有意な識別が可能である。

顔画像識別の際には静止した顔を処理すればよいので、応用によっては顔を固定するという制約も導入可能である。しかし、表情画像の場合、動画像(複数画像)処理が必要となり、さらに一般的な応用としては自然な状態での入力が前提となる。そこで、次のような困難な課題を解決する必要がある。

課題1: 顔全体の揺れを補正する必要があるにもかかわらず、補正の手がかりとなる造作はすべて変形している。

課題2: 表情や感情は一瞬の状態の解釈ではなく、時間情報を含んだパターンであるが、FACSのような定性的記述モデルしか与えられていない。

#### 4. 最近の技術動向

Black & Yacoob<sup>10)</sup>は、顔と造作を平面モデルで仮定して、画像の動きを3種の多項式で表現することで、図2のように、顔全体の剛体運動と、口と眉のゆ

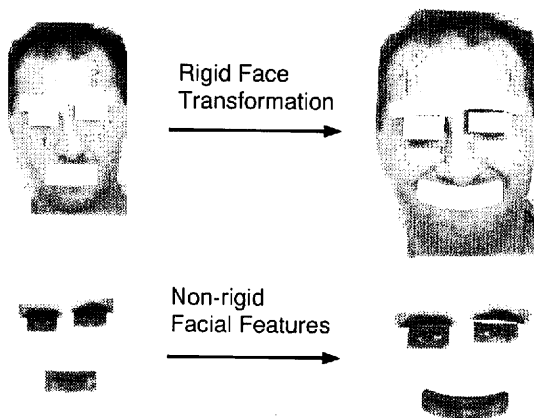


図2 顔全体の剛体運動と造作の変形の同時抽出(写真提供: Michael Black氏, Xerox PARC)

がみによる柔物体運動を抽出している。抽出された変形はパラメータ値から「口が開いている」などの中間表現への変換と、その「幸福」などの上位表現への変換論理を経て表情認識結果を得ている。このように、顔全体の動きの補正ができるようになると、実世界での応用範囲が大きく広がる。実例としてTV番組のクリップからも表情認識の実験を行っている。

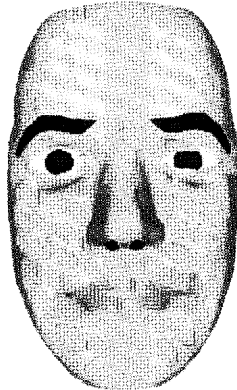
Essa & Pentland<sup>11)</sup>はオプティカルフローをそのまま用いずに、顔のメッシュモデル構造に基づきフローの最適推定をして、より外乱に強い安定した変形特徴パターンを得ている。さらにそれを顔の動きエネルギーの時空間分布(図3を参照)としてモデル記述して、パターンマッチングにより認識を行っている。また、静止画ベースの記述であるFACSのダイナミックモデルへの拡張も提案している。

時系列パターンの認識手法として有効な隠れマルコフモデル(HMM: Hidden Markov Model)を利用することも試みられている。対象が顔表情の場合には各表情に無表情という共通の起点があるため、シンボル列の設計に注意が必要である。坂口ら<sup>12)</sup>はシンボル化の際に用いるカテゴリー別ベクトル量子化を改良して良好な認識性能を得た。

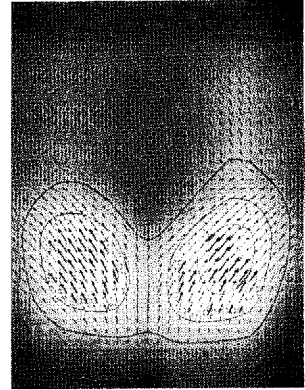
表情認識の性能は最終的には認識率ということになる。上記の報告はいずれも、喜び、怒り、驚き、悲しみ、嫌悪、恐れ、の6基本感情表情と瞬き、眉あげの表情のうち4~6種の表情データの認識を行って、認識率を報告している。実験条件が違うため単純に比較することは誤解を招くおそれが大いにあるが、現状を知るうえで参考になると思われるので、各論文の著者が主張している平均認識率を表1にまとめておく。



(a) 入力表情



(b) モデル表情



(c) 動きエネルギー

図3 オプティカルフロー特徴と顔のモデルを使った表情動作のエネルギー表現  
(写真提供: Irfan Essa 氏, ジョージア工科大)

表1 各手法の表情認識率 (注意: 条件は同一でない)

手法	特徴	カテゴリー数	データ数	平均認識率
Mase 91 <sup>9)</sup>	フロ・特徴とk-最近傍識別	4	22	86%
Black ら 95 <sup>10)</sup>	剛性動き補正と変形記述	6	70	92.0%
坂口ら 95 <sup>12)</sup>	HMM+VQ	4	48	93.7%
Essa ら 95 <sup>11)</sup>	動きエネルギー分布	5	52	98.0%

## 5. む す び

表情認識研究におけるもうひとつの大きな課題は、「自然な表情」認識である。ここでとりあげた論文の実験はほとんど人為的な表情を実験データとしており、強調されたものであることを理解しておかなければならない。自然な表情とは何か、興味や意図や情動との関連の手がかりはどこか、さらに調査する必要がある。

ところで、経験的に「笑う」という顔表情は作りやすいが「怒り」や「悲しみ」は困難である。面白いことに、声の表情はそれが逆転するようだ。マルチモーダルインタフェースにおける表情メディアの機能は他のメディアとの統合なしには考えられない。

(1997年5月15日受付)

## 【参 考 文 献】

- 1) 間瀬 健二: “動画画像処理を用いた新しいマンマシンインタフェースの研究”. 名古屋大学学位論文 (Mar. 1992)

- 2) <http://mambo.ucsc.edu/psl/fanl.html> (表情分析に関する home page)
- 3) P. Ekman and W. V. Friesen: “The Facial Action Coding System”. Consulting Psychologists Press, Inc., San Francisco, CA (1978)
- 4) K. Waters: “A Muscle Model for Animating Three-dimensional Facial Expression”, *Computer Graphics*, 21, 4, pp. 17-24 (1987)
- 5) 相澤, 原島, 斉藤: “構造モデルを用いた画像の分析合成符号化方式”, *信学論*, J72-B-1, 3, pp. 200-207 (Mar. 1989)
- 6) 福井: “モーションキャプチャ”. 映像情報メディア学会誌「人体と顔の画像処理」小特集 (Aug. 1997)
- 7) 海老原, 大谷, 岸野: “臨場感通信会議のための実時間表情検出”, *テレビ誌*, 50, 10, pp. 1497-1506 (Oct. 1996)
- 8) K. Mase, “Recognition of Facial Expression from Optical Flow”, *IEICE transactions (E)*, E-74, 10, pp. 3474-3483 (Oct. 1991)
- 9) P. Ekman, T. S. Huang, T. J. Sejnowski and J. C. Hager (editors): “Final Report to NSF of the Planning Workshop on Facial Expression Understanding” (July. 1992)
- 10) M. J. Black and Y. Yacoob: “Recognizing Facial Expressions under Rigid and Non-Rigid Facial Motions.” *Int. Workshop on Automatic Face- and Gesture-Recognition*, Zurich, pp. 12-17 (Sep. 1995)
- 11) I. A. Essa and A. P. Pentland: “Coding, Analysis, Interpretation, and Recognition of Facial Expressions”, *Technical Report 325*, Perceptual Computing Section, MIT Media Laboratory (Apr. 1995)
- 12) 坂口, 大谷, 岸野: “隠れマルコフモデルによる顔動画像からの表情認識”, *テレビ誌* 49, 8, pp. 1060-1067 (Aug. 1995)



間瀬 健二 1981年, 名古屋大学工学部大学院修士(情報)課程修了。同年, NTT入社。1988~1989年, 米国MITメディア研究所客員研究員。現在, (株)ATR知能映像通信研究所第二研究室長, コンピュータグラフィックス, 画像処理とそのヒューマンインタフェースへの応用等の研究に従事。工学博士。