

## ユーザモデルを用いた感情通信システム

倉石英俊 柴田義孝

東洋大学 工学部 情報工学科

{kuraishi,shibata}@yosemite.sb.cs.toyo.ac.jp

民族性や年齢の相違による感情の認識不足により、コミュニケーションがスムーズに行なわれなかったために、相互理解にしばしば支障をきたすことがある。このような問題を解決するための方法として、表情の個人差を考慮し、正確な感情を伝えることができる通信システムが有効である。本研究では、ある感情に対する表情の個人差を個人モデルとしてモデル化し、表情からの正確な感情抽出および抽出した感情を相手に正確に伝えるため表情の合成を行なうことにより、より正確な感情の伝達を可能とする、感情通信システムを提案する。本稿では、感情通信システムおよび個人モデル、表情のモデル化の方法について述べ、その有効性を考察する。

## Feeling Communication System Using User Model

Hidetoshi Kuraishi, Yoshitaka Shibata

Department of Information and Computer Sciences

Toyo University

{kuraishi,shibata}@yosemite.sb.cs.toyo.ac.jp

Recognition insufficiency of feeling due to the differences of age and racial characters often interferes with mutual understanding between human beings. In this paper, a feeling communication system which enables to convey more accurate feelings by modeling of individual differences of facial expression as "Individual Models" is proposed. By using this communication system, real feeling from one user's facial expression can be extracted, transmitted to the another user, and eventually a suitable facial expression is synthesized so as to correctly understand his feeling to the another user.

### 1 はじめに

我々日本人は他国民族に比べ感情の表現力に乏しいと言われる。また他国民族は我々から見て表現が豊かであるため、互いに表情から感情を理解する場合に認識のズレが生じ、国際的なコミュニケーションの場などにおいて相互理解にしばしば支障をきたすことがある。

一方、ネットワークの高速化やマイクロプロセッサの高性能化などにより、高精細な画像通信による実際の人間の表情に忠実な顔画像を転送、再生することも可能となりつつある。ま

た、人間の顔の表情に注目して顔の筋肉の動きをパターン化し、特徴を転送する顔画像通信により超低ビットレート転送を実現する研究も盛んに行なわれている [1][2] が、表情には個人により差があり、同じ感情に対してもすべての人間に同様の表情を示すことは現実的ではなく、これらの研究はそのような表情の個人差を考慮するものではなかった。

そこで本研究では、通常のビデオ通信システムと感情通信システムを組み合わせることにより、より相互理解を深めることができるシステ

ムを考察した。

本研究で提案する感情通信システムは、ある表情における筋肉の動きの特徴をパターン化し、パターンの個人差を分析側の個人モデル、また相手が正確に理解できる表情における筋肉の動きをパターン化し、パターンの個人差を合成側の個人モデルとして表情の分析合成システムに組み込み、表情の一般化とパターンマッチングによる個人の正確な感情の抽出および表情の個性化とコンピュータグラフィックス(CG)による相手に正確な感情を伝える表情の合成を行なうことにより正確な感情を伝えることができるシステムであり、初対面の人間同士のコミュニケーションやフォーマルな場でのコミュニケーションなどの相互理解が必要とされる場合において利用されることを前提とする。

## 2 感情と表情と個人モデル

感情を表現する手段として、一般的にジェスチャー [3]、言葉、表情の3つがあげられるが、我々は表情に着目する。

本システムにおいて伝達の対象とする感情は、現段階では、喜び、怒り、悲しみ、驚き、恐れ、嫌悪の基本的な6通りであり、各感情においてその度合を考慮する。各感情に応じて表情が存在し、感情の度合は表情の変化の大きさにより考慮する(4節参照)。

また個人の表情とは、ある感情に対する個人の表情であり、一般的表情とは、ある感情に対する大多数の人間に共通に見られる表情であると定義する。

感情⇔一般的表情⇔個人の表情の変換を写像変換と考え、本研究における感情と表情との関係を定義する。図1は、感情 $F$ が与えられたとき、写像 $g$ によって一般的表情 $E_{gen}$ が得られ、一般的表情から写像 $h_A$ により個人の表情 $E_{per,A}$ ( $A$ : Analysis) が得られ、また写像 $h_S$ により個人が相手の感情を理解できる表情 $E_{per,S}$ ( $S$ : Synthesis) が得られることを示す。

$$g(F) = E_{gen}$$

$$h_A^{-1}(E_{gen}) = E_{per,A}, h_{per,S}(E_{gen}) = E_S$$

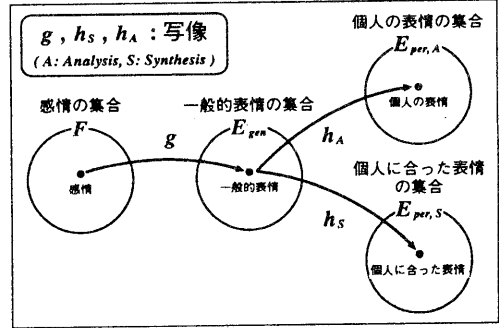


図1. 感情と表情との関係

分析側の個人モデルとは、個人の正確な感情を抽出するために個人の表情を一般的表情に変換する(一般化)モデルであり、写像 $h_A^{-1}$ において適用される。

合成側の個人モデルとは、相手の感情の正確な理解を可能とするために一般的表情を個人に合った表情に変換する(個性化)モデルであり、写像 $h_S$ において適用される。

## 3 個人モデルを用いた感情通信

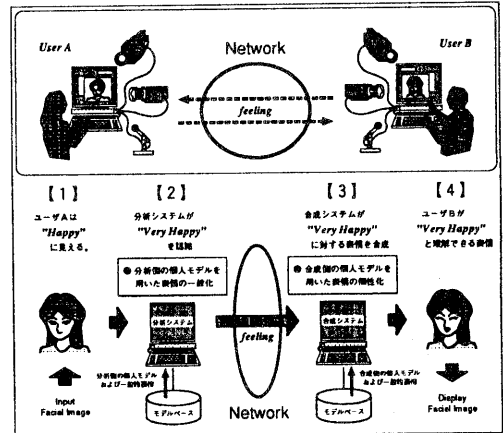


図2. 個人モデルを用いた感情通信

図2は、個人モデルを用いた感情通信の流れを、表現力に乏しいユーザAと感情をやや控え目に理解するユーザBがネットワークを介してテレビ電話のような通信を行なっているという例を用いて示している。

- 1) ユーザAは本当は“非常に嬉しい”のだが、表現力が乏しいために表情は一般的に見て

“ただ嬉しい”を表現しているように見える。カメラはユーザAの“ただ嬉しい”の表情を忠実に取り込む。

- 2) 分析システムは、ユーザAの“ただ嬉しい”の表情をユーザAの分析側の個人モデルを用いて一般化し、感情を判別することにより、ユーザAの正確な感情 (“非常に嬉しい”)を抽出する。そして感情情報 (感情、感情の度合)を合成側に転送する。
- 3) 合成システムは、受信したユーザAの感情 (“非常に嬉しい”)に対する一般的表情をユーザBの合成側の個人モデルを用いて個性化し、ユーザBがユーザAの感情を正確に理解できる表情 (少しオーバーな表情)を合成する。
- 4) ユーザBは表示されたユーザAの表情から、ユーザAが“非常に嬉しい”ことを理解できる。

ここで、感情の表現は必ずしも元の人間の顔に忠実な表情によって行なう必要はなく、アニメーションや言葉、文字などでもよいと考える。本稿では、感情を表現するものとして原顔画像による表情を用いたときの感情通信システムを説明する。

#### 4 分析側のモデル

表情の分析し、正確な感情を抽出するために、3次元ワイヤフレームモデル (3-D WFM) に加えて以下のモデルを用いる。

- 感情と一般的表情との対応を示すモデル
- 一般的表情のモデル
- (分析側の) 個人モデル

感情と一般的表情との対応として、1つの感情に1種類の一般的表情のモデルが対応すると定義する (喜び→微笑み、笑顔)。また分析側での表情は、3次元データを基にし、これを平面のみの表示のために2次元データとして扱われる。

#### 4.1 表情のモデル化

分析側で扱われる表情は、顔上に設定された特徴点の動きの組合せである表情パターンとして2次元マトリクスでモデル化される。

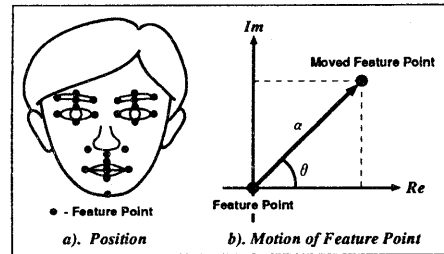


図3. 特徴点

特徴点とは、表情を作る際に動きが顕著に現れる点であり、現段階では図3 a)の26点を設定している [1]。特徴点の動きは図3 b)のように、移動量 $\alpha$ と移動角度 $\theta$ を用いて $\alpha e^{j\theta}$ という2次元ベクトルとして表される。

個人の表情パターン  $E_{per}$ は、特徴点の動きの組合せで表される。

$$\begin{aligned} E_{per} &= [\alpha_1^{j\theta_1} \alpha_2^{j\theta_2} \dots \alpha_{26}^{j\theta_{26}}]^T \\ &= [a_1 + jb_1 \dots a_{26} + jb_{26}]^T \quad (1) \end{aligned}$$

また複数の人間の表情パターンにおいて、式(1)の実数部 ( $a_i$ )と虚数部 ( $b_i$ )をそれぞれ平均した値 ( $\bar{a}_i, \bar{b}_i$ )を用いて得られるパターンを一般的表情パターン ( $E_{gen}^f, 1 \leq f \leq 7$ :感情のインデクス)と呼び、次式で表す。

$$E_{gen}^f = [\bar{a}_{f,1} + j\bar{b}_{f,1} \dots \bar{a}_{f,26} + j\bar{b}_{f,26}]^T \quad (2)$$

一般的表情パターンは、6通りの感情に対する6つの表情に無表情を加えた計7つの各表情に対して1つずつ存在する。無表情に対する一般的表情パターンのすべての要素 (特徴点の動き)は0である。

また感情の度合を決定する要素の一つとして表情の大きさ ( $A$ : Amplitude)を定義する。個人の表情の大きさは式(3)で与えられ、一般的表情の大きさは式(4)で求められる。

$$A_{per} = \frac{1}{26} \sum_{i=1}^{26} \sqrt{a_i^2 + b_i^2} \quad (3)$$

$$A_{gen} = \frac{1}{26} \sum_{i=1}^{26} \sqrt{\bar{a}_i^2 + \bar{b}_i^2} \quad (4)$$

## 4.2 分析側の個人モデル

ある感情に対する一般的な表情変化に対する筋肉の動きとユーザ特有の表情変化に対する筋肉の動きとの違いを2次元マトリクスでモデル化し、これを分析側の個人モデルとする。分析側の個人モデルを用いて個人の表情を一般化し、分析を行なうことにより、個人の正確な感情を取り出すことができると考える。

一般的表情パターン  $E_{gen}$ 、個人の表情パターン  $E_{per}$ 、分析側の個人モデル  $M_A$  との関係は次式のように定義する。

$$E_{gen} = M_A E_{per} \quad (5)$$

$$M_A = \begin{bmatrix} m_{1,1} & \cdots & m_{1,26} \\ \vdots & \ddots & \vdots \\ m_{26,1} & \cdots & m_{26,26} \end{bmatrix} \quad (6)$$

式(5),(6)と前に示した式(1),(2)から、分析ユーザモデルの要素  $m_{i,i}$  は次式で表される。

$$m_{i,i} = \lambda_{i,i} e^{j\omega_{i,i}} \quad (7)$$

$$\lambda_{i,i} = \sqrt{\bar{a}_i^2 + \bar{b}_i^2} / \sqrt{a_i^2 + b_i^2}$$

$$\omega_{i,i} = \tan^{-1}(\bar{b}_i/\bar{a}_i) - \tan^{-1}(b_i/a_i)$$

分析側の個人モデルは一般的パターンと同じく各表情について1つ(計7つ)存在する。無表情に対する個人モデルは単位行列とする。

## 5 合成側のモデル

感情を正確に伝える表情を合成するために、WFMを除き以下のモデルを用いる。

- 感情と一般的表情との対応を示すモデル
- 各表情に対応する筋肉の動作モデル
- (合成側の)個人モデル

感情と一般的表情との対応は分析側と同じく、感情に対して1つの一般的表情が存在する。

## 5.1 表情のモデル化

合成で用いられる表情は、顔器官の基本動作(Action Unit: AU, 図4 a)参照)[4]の組合せで表される。そして各AUはキーノードの3次元移動ベクトルの集合である動作マトリクスで表される。

各表情は、例として図4 b)のようなAUの組合せをとる[5]。

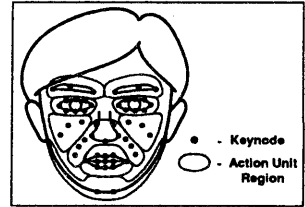
キーノードとは、AUの動作の中心となるWFMのノードであり、図4 c)の黒点の位置に設定している。AUの境界は、各AUを実行する際、そのAUの境界内のノードがキーノードの動きに従って移動することを示す。

AU番号	AU名	AU番号	AU名
1	眉の内側を上げる	10	上唇を上げる
2	眉の外側を上げる	12	唇両端を引き上げる
4	唇を下げる	15	唇両端を下げる
5	上まぶたを上げる	17	下顎を上げる
6	頬を持ち上げる	20	唇両端を横に引く
7	まぶたを緊張させる	25	顎を下げずに唇を開く
9	鼻にしわをよせる	26	顎を下げて唇を開く

a). AUの例

表情	AU番号(例)
喜び	6+12+26
怒り	4+5+7+10+26
愁しみ	1+4+15
驚き	1+2+5+26
恐れ	1+2+4+5+7+20+26
厭悪	4+9+17

b). AUによる表情のモデル化



c). キーノードおよびAUの領域

図4. AUと表情とキーノード

AU番号  $i$  のAU( $AU_i$ )の個人の動作マトリクス  $V_{per}^i$  は  $AU_i$ の境界内のキーノード ( $n_i$ 個,  $1 \leq k \leq n_i$ ) の3次元移動ベクトルの集合であり、次式のように表される。

$$V_{per}^i = [v_{i,1} \cdots v_{i,n_i}]^T, \quad v_{i,k} = \{x_{i,k}, y_{i,k}, z_{i,k}\}^T \quad (8)$$

一般的表情の動作マトリクス  $V_{gen}^i$  を、一般的な人間の、キーノードの移動ベクトルの集合として次式のように表す。

$$V_{gen}^i = [\bar{v}_{i,1} \cdots \bar{v}_{i,n_i}]^T, \quad v_{i,k} = \{\bar{x}_{i,k}, \bar{y}_{i,k}, \bar{z}_{i,k}\}^T \quad (9)$$

## 5.2 合成側の個人モデル

ある感情に対して、一般的な人間がその感情を理解できる表情と個人が理解できる表情との違いをマトリクス形式でモデル化し、これを合成側の個人モデルとする。合成側の個人モデルを用いることにより個人が相手の感情を正確に理解できる表情を合成することができると考える。

$AU_i$ (境界内のキーノードの数: $n_i$ )の一般的動作マトリクス  $V_{gen}^i$ 、個人の動作マトリクス  $V_{per}^i$ 、 $AU_i$ に対する合成側の個人モデル  $M_S^i$ の間の関係を次式のように定義する。

$$V_{per}^i = M_S^i V_{gen}^i \quad (10)$$

式(10)と前に示した式(8),(9)から、合成側の個人モデル  $M_S^i$ は次式のように表される。

$$M_S^i = [M'_{i,1} \cdots M'_{i,n_i}] \quad (11)$$

$$M'_{i,k} = \begin{bmatrix} r_{x_{i,k}} & 0 & 0 \\ 0 & r_{y_{i,k}} & 0 \\ 0 & 0 & r_{z_{i,k}} \end{bmatrix} \quad (12)$$

$$\begin{aligned} r_{x_{i,k}} &= x_{i,k} / \overline{x_{i,k}}, & r_{y_{i,k}} &= y_{i,k} / \overline{y_{i,k}} \\ r_{z_{i,k}} &= z_{i,k} / \overline{z_{i,k}}, & (1 \leq k \leq n_i) \end{aligned}$$

合成側の個人モデルは各 AU に対して1つ存在している。

## 6 感情通信システム

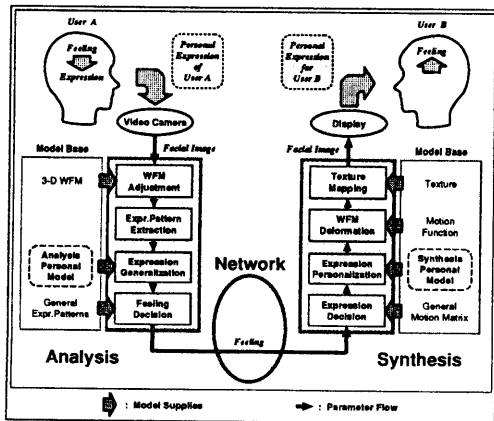


図5. システムの構造

処理モジュールに沿って処理の流れを説明する。分析側では各感情に対する一般的表情パターンと分析側の個人モデル、合成側では一般的動作マトリクスと合成側の個人モデルがあらかじめ登録されており、それらのモデルはモデルベースに格納されている。

- 1) カメラが顔画像を取り込む。
- 2) **WFM Adjustment** : WFM を顔画像に整合し、特徴点の動き検出を行なう。
- 3) **Expression Pattern Extraction** : 特徴点の動きを用いて個人の表情パターンを作成する。
- 4) **Expression Generalization** : 分析側の個人モデル  $M_A$ を用いて、個人の表情パターンを  $E_{per}$ を次式で一般化し、一般化した表情パターン  $E'_{gen}$ および表情の大きさ  $A'_{gen}$ を得る。

$$E'_{gen} = M_A E_{per} = [p_1 + jq_1 \cdots p_{26} + jq_{26}]^T \quad (13)$$

$$A'_{gen} = \frac{1}{26} \sum_{i=1}^{26} \sqrt{p_i^2 + q_i^2} \quad (14)$$

- 5) **Feeling Decision** : 一般化した表情パターン(式(13))とあらかじめ登録されている一般的表情パターン  $E_{gen}^f$ (式(2)、 $1 \leq f \leq 7$ :感情のインデクス)を比較し、その誤差を次式により求め、誤差が最小となる一般的表情パターンに対する感情を抽出する。

$$Err_f = \sum_{i=1}^{26} \{(Rerr_{f,i} + Ierr_{f,i}) \times W_{f,i}\}$$

$$Rerr_{f,i} = (\overline{a_{f,i}} - K_f p_i)^2$$

$$Ierr_{f,i} = (\overline{b_{f,i}} - K_f q_i)^2 \quad (15)$$

$K_f$ は感情の度合を表す係数であり、一般化した表情パターンの大きさ  $A'_{gen}$ (式(14))と一般的表情パターンの大きさ  $A_{gen}^f$ (4節式(4))を用いて次式で求められる。

$$K_f = \begin{cases} A_{gen}^f / A'_{gen} \\ 1 \end{cases} \quad (\text{無表情のとき}) \quad (16)$$

無表情には感情の度合がないため  $K_f$  は 1 とする。また  $W_{f,i}$  は特徴点により異なる重みである。ある特徴点がユーザ特有の動きであればその特徴点の重みは大きな値をとる。

- 6) 感情情報として、“喜び”、“驚き”といった感情を表す単語と感情の度合を表す係数 ( $K_f$ ) を合成側へ転送する。
- 7) **Expression Decision** : 感情を表す単語および感情の度合を表す係数  $K_f$  を受信し、受信した感情に対する表情を構成する AU の AU 番号を得、その AU 番号に従って AU の一般的動作マトリクスを得る。
- 8) **Expression Personalization** :  $AU_i$  の一般的動作マトリクスを  $V_{gen}^i$ 、この AU に対応する合成側の個人モデルを  $M_S^i$ 、個人の動作マトリクスを  $V_{per}^i$  とすると、 $V_{per}^i$  は次式により求められる。

$$V_{per}^i = K_f M_S^i V_{gen}^i \quad (17)$$

- 9) **WFM Deformation** : 動作マトリクスに従って、WFM のノードを移動させる。
- 10) **Texture Mapping** : テクスチャをはる。
- 11) 表示。

## 7 実装上の問題点

本システムの実現上の問題点を上げる。

### 分析側

- 特徴点の動き検出の自動化
- 表情パターン正確性および分析側の個人モデルの十分な特徴の考慮

### 合成側

- 一般的動作マトリクスの正確性および合成側の個人モデルの十分な特徴の考慮
- 時間概念を加えた滑らかな表情変化の実現

### システム全体

- 適切な感情の度合を考慮する分析合成方法
- 表情だけでは表現できない感情の表現方法: ジェスチャーや言葉との組合せによる表現
- 評価方法として、様々な人間 (外国人、年齢の違う人間等) 同士のコミュニケーションがスムーズに行なえるか。

## 8 まとめ

これまで感情通信システムの提案、個人モデルおよび表情のモデル化方法について述べた。

本システムは感情の伝達だけでなく、実際の表情とそれに対する感情をユーザに伝えることによりコミュニケーションにおける相互理解をより正確にすると考えられる。また本システムを用いてコンピュータに人間の感情を理解させることにより、より人間に近いコンピュータとの対話を行なうアプリケーションを作成することも可能であると考えられる。現段階では一般的表情パターンの作成、分析側の個人モデルの作成を行ない、それらを用いた感情抽出、また CG による WFM の作成、表示、6 通りの感情に対する AU の動作マトリクスを用いた表情変化を行なっている。今後は、これまでに述べた方法論を基本にして、問題点を解決し、実装し、個人モデルの有効性の評価を中心に本感情通信システムの機能評価を行なう予定である。

## 参考文献

- [1] K.Aizawa, H.Harashima, T.Saito, "Model-based analysis synthesis image coding (MBASIC) system for person's face", *Signal Process., Image Com.*, vol.1, no.2, pp.139-152 (Oct. 1989)
- [2] C.S.Choi, H.Harashima, T.Takebe, "Analysis and synthesis of facial expressions in knowledge-based coding of facial image sequences", *IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP-91)*, 29M9.7 (May 1991)
- [3] T.Inoue, K.Okada, Y.Matsusita, "Using gestural animations to convey the feelings in asynchronous network communication", ICOIN-J005
- [4] 崔昌石, 原島博, 武部幹, "顔の3次元モデルに基づく表情の記述と合成", *信学論 (A)*, J73-A, 7, pp.1270-1280 (1990-07)
- [5] P.Ekman, W.V.Friesen, "Facial Action Coding System", Consulting Psychologists Press Inc., 577 College Arc, Palo, Alto, 1977. Jul.1994.