

感情通信システムにおける表情の分析および合成法の評価

坂本祐之 松尾省吾 柴田義孝

東洋大学 工学部 情報工学科

{sakamoto,matsuo,shibata}@sb.cs.toyo.ac.jp

通信に限らず我々のコミュニケーションにおいて、互いの感情を正確に理解することは大変重要な要素である。筆者等はこれまで、感情表現、認識の個人差のモデルをもちいて表情分析、合成を行なうことで、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している。このためには、分析側の個人モデルとして表情により感情を表現するときのユーザ個人のもつ特徴を表すモデルと、合成側の個人モデルとして表情によって感情を理解するときのユーザのもつ特徴を表すモデルを用意する。これらを用いる事によって、より正確な感情をユーザの表情から抽出することと、より正確な感情をユーザに伝達する表情を合成する事が可能となる。本稿では、感情通信システムにおける表情の分析法の評価と視線と顔の向きをもつ表情の合成について述べる。

Evaluation of Facial Expression Analysis/Synthesis for Feeling Communication System

Hiroyuki Sakamoto, Syougo Matsuo and Yoshitaka Shibata

Department of Information and Computer Sciences

Toyo University

{sakamoto,matsuo,shibata}@sb.cs.toyo.ac.jp

Recognition of accurate feeling each other is very important for mutual understanding for human beings. We proposed a feeling communication system which enable to convey more accurate feeling by considering the difference of facial expression and the difference of feeling recognition from one facial expression using individual models. In our system, an analytical individual model is defined as a model which shows features of the individual expression. Beside, an synthetical individual model is defined as a model which shows features of facial expression which one person can understand an fellow's feeling correctly. Using individual model made possible recognize and convey accurate feeling. In this paper, we describe about Evaluation of feeling recognition method and Facial Expression synthesis which deals with face and eye directions.

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いわれる。また、他の民族の中には我々から見て感情表現が豊かであるために感情がオーバーに感じられる場合も多い。このような感情表現の違いは、互いの感情を理解しようとする場合に正しい感情の認識を阻害することがある。我々はこのような問題を解決するために、個人モデルに基づく表情の

分析および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している [1]。

本システムは、分析側の個人モデルを用いてユーザ個人の表情からそのユーザの特徴を抽出し(表情の一般化)、感情を判別することで、より正確なユーザの感情を取り出す機能、および合成側の個人モデルを用いて一般的表情にユーザの見方を考慮した特徴を加え(表情の個性化)、その表情をコンピュータ

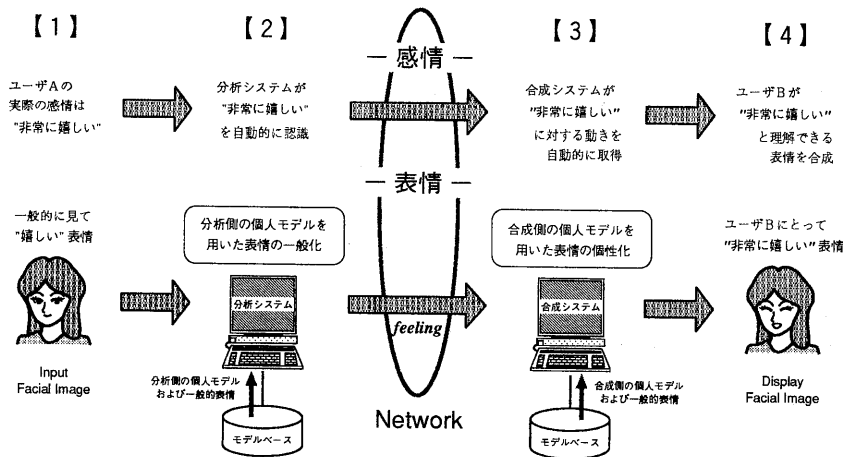


図1. 個人モデルを用いた感情通信

グラフィックス (CG) を用いて合成、表示する機能により正確な感情の伝達を行なうシステムであり、通常のビデオ通信システムと組み合わせることにより、初対面の人間同士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も円滑なコミュニケーションを行なう事ができる。

本稿では、分析システムにおける表情認識の自動化のためのプロトタイプの実装と、合成システムにおけるそれまでのCGによる表情の合成に視線と顔の向きを要素を追加することで感情の表現力を向上させる試みについて述べる。

2 感情通信システム

感情通信システムは、基本的な6つの感情、“喜び” “怒り” “悲しみ” “恐れ” “嫌悪” “驚き”) とそれぞれの感情の度合を組み合わせて感情を伝達する。

感情を表現する方法としては、バーバルな手段とノンバーバルな手段に大別されるが、特に顔の表情による感情伝達の比重が高い [2] ことから本研究では、顔による表情の合成を行なう方法を選択した。感情通信システムは、図2のように1対1で通信をおこなっているときに感情を控え目に表現するユーザAからユーザBにたいして感情を伝達する場合、図1の様にして感情を正確に伝達する。

- 1) ユーザAの感情は“非常に嬉しい”のであるがユーザAは控え目な表情をするため、“少し

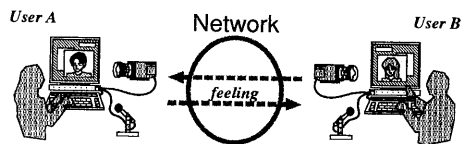


図2. 1対1通信

嬉しい”ように見える。

- 2) 分析システムは、ユーザAが平均的な人間よりも控え目な表情をすることを表す分析側の個人モデルをもちいることにより表情から“非常に嬉しい”感情を正確に抽出し、合成側に伝達する。
- 3) 合成システムは、ユーザBが理解できる表情と平均的な人間が理解できる表情との違いを表す合成側の個人モデルをもちいることで、ユーザBが“非常に嬉しい”と理解できる表情を合成する。
- 4) ユーザBに対して表情による感情の認識の個人差を考慮した表情が提供される。ユーザBはこれを見てユーザAの“非常に嬉しい”感情を正確に理解する。

感情通信システムにおいて、このような感情表現および認識の個人差の考慮を行なうためには、平均的な表情とそれに対する個人差を正しく定義する必要がある。

3 分析システムにおける感情認識

人間の表情に注目した画像処理の研究には、オペティカルフローをもちいて表情の変化から個人識別を行なうもの [3]、赤外線画像を用いて表情の認識を行なうもの、ポテンシャルネットとKL展開を組み合わせて表情の認識を行なうもの [4] などがある。本研究では表情における微妙な個人差などを認識するため、分析システムでは、表情を認識するために図3のように表情をつくる顔の動きが顕著に現われる26点を特徴点として設定し、表情をこれら特徴点の動きとしてとらえることで、より詳細な表情の認識を試みている。表情は、特徴点の移動量 α_i と、移動角度 θ_i から次式の $Eanaly$ として表現される。

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

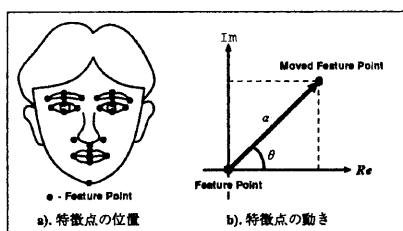


図3. 分析システムにおける特徴点

図4は、分析システムにおいてビデオカメラから感情を抽出し、ユーザの感情を抽出するまでの処理の流れを示している。

1) ビデオカメラによる顔画像の取り込み

ビデオカメラなどの入力装置より、ユーザの顔画像を得る。

2) 特徴点の位置探索

ビデオカメラからとりこまれた特徴点の位置を求めるための方法として、本研究では顔の各部位の位置から相対的に特徴点の位置を求める方法を考え、このための画像処理法について検討を進めている。この方法では、特徴点の位置を顔の部位の位置から求めるために、あらかじめ必要なルールを用意しておかなければならない。また、たとえば口唇周辺と目の周辺では、

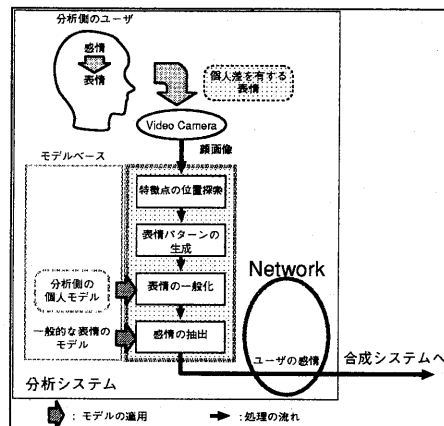


図4. 分析システムの処理

特徴点とそれ以外の部分を判別する最適な方法は異なることが予想されるため、特徴点毎に適した複数の探索法を検討しなければならない。

3) 表情パターンの生成

特徴点の位置から特徴点ごとの移動ベクトルが得られる。この移動ベクトルから生成される式(1)の形式の表情パターン $Eanaly^{indiv}$ が得られる。

4) 表情の一般化

表情を各特徴点毎の移動ベクトルとして表現する場合には、個人差はこれらの差分として表現することが可能である。ある感情 f の個人モデル M_f^{analy} と個人差を含む表情パターン $Eanaly^{indiv}$ 、一般的な表情 $Eanaly_f^{gen}$ の間には、式(2)の関係が成り立つ。

$$Eanaly_f^{gen} = M_f^{analy} \cdot Eanaly^{indiv} \quad (2)$$

5) 感情の抽出

一般化されたユーザの表情と、各感情に対応した表情のモデルの比較を行ない、ユーザの感情を判別する。具体的には、式(2)の両辺の差が最小となる感情を選択することで行なわれる。

4 分析システムのプロトタイプ

分析システムを実装するためには、ビデオカメラより取り込まれた画像から、まず特徴点の位置を特定し、これの動きを追跡することが必要となる。特

特徴点の位置を特定する方法としては、顔を構成する目、口等の部品の位置と一定のルールからその位置を決める方法が考えられる。またこのような顔の各部分の位置を探索する方法としては、濃淡画像から口唇領域を抽出する方法 [5]、色相の違いから目及び瞳の位置を検出する方法 [6] 等があるが、本研究のプロトタイプでは、まず単純なテンプレートマッチングで特徴点の探索、追跡を試み、その結果をもとに今後の方針を検討した。図5は、プロトタイプであらかじめマークをつけた3つの特徴点を探索した例である。

この時、ほぼ正確に追跡が行なえ、適当な明るさを提供する光源がある環境下ではマークなしでも特徴点の追跡が行なえた。しかし、周囲が暗い場合には、マークをつけていた場合でも特徴点を正確に追跡することはできず結果は非常に不安定になった。このことから今後は光源などの環境に影響を受けにくい方法として色相の違いを利用する方法や、特徴点の位置によってはその追跡にオプティカルフローの生成に広く利用されている時空間勾配法などの方法の導入の検討が必要であると考えている。

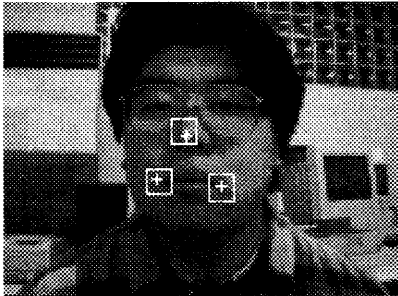


図5. プロトタイプによる特徴点の探索

5 合成システムにおける表情合成

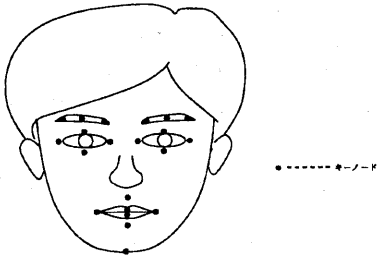


図6. 合成システムにおけるキーノードの配置

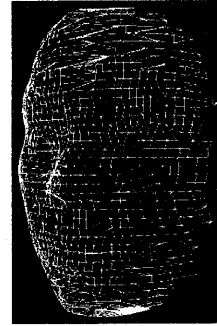


図7. WFM

顔の表情を合成する研究としては、モデルベース符号化技術として顔の筋肉の動きをパターン化し、その特徴を転送する顔画像通信がある [7]。しかし、これらの研究においては、超低ビットレート転送を可能とするために元の人間の表情を忠実に再現するための技術が主体であるが、個人によってその感情に対する表情が異なるので、必ずしも相手に正しくその感情を伝達できるとは限らない。

そこで、合成システムは顔の表情をCGを利用して合成し、感情を表現するための顔画像には本人の原顔画像を使用する。また表情の合成においては表情のみならず、視線と顔の向きも考慮することにより、感情の表現力の向上を図っている。

顔の表情変化を計算機上で表現するために、図6のように顔に動きを制御するための点「キーノード」を設定している。本システムで扱う顔モデルは、WFM(ワイヤーフレームモデル)とよぶ図7のような三次元上のノードの集合であり、キーノードはこれらの動きを制御するために用いられている。このキーノードをもちいることで、合成側において表情は、対応する感情を f とおくと、キーノード i の移動ベクトル V_i を用いて次式のように表現される。

$$E_{synthf} = [V_1, \dots, V_{21}]^T \quad (3)$$

図8は、合成システムが、分析システムより相手の感情を受け取り、ユーザに合成された表情を提供するまでの処理の流れを示している。

1) 表情の取得

分析システムからは、感情とその度合の情報

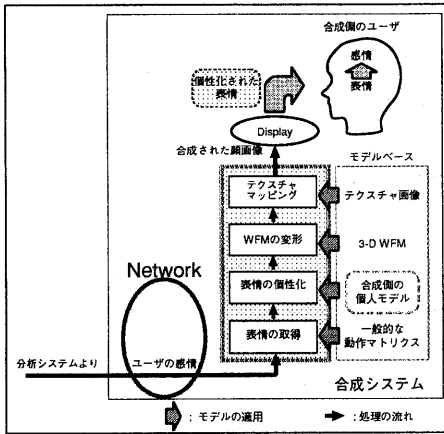


図8. 合成システムの処理の流れ

が送られて来る。合成システムは、感情から式(3)で表される対応した表情のモデルを得る。

2) 表情の個性化

最初に得られる表情のモデルは、ユーザの感情認識の個人差を考慮するものではない。そこで感情認識の個人差のモデルをもちいて表情を個性化し、ユーザが正しく感情を認識できる表情のモデルを生成する。合成側の個人モデルはキーノードの移動ベクトルの差分として表現され、個性化された表情 $E_{synth_f}^{indiv}$ は、一般的な表情 $E_{synth_f}^{gen}$ とユーザの個人モデル M_f^{synth} をもちいて次式で与えられる。

$$E_{synth_f}^{indiv} = M_f^{synth} \cdot E_{synth_f}^{gen} \quad (4)$$

3) WFMの変形

WFMを変形し表情を合成する。WFMを実際に変形するには大きく分けて2つの処理が行なわれる。

a. キーノードの移動

キーノードは顔の動きの特徴、とりわけ顔の各部位の動きの特徴を表す位置に配置されている。キーノードの移動ベクトルは、式(4)で与えられており、これを用いて顔のおおまかな変形を行なう。

b. 他のノードの移動

キーノードの動きを顔全体に反映させる。各ノードにはあらかじめその位置に基づいて、キーノードの動きを元に自分の動きを決定するためのルールが与えられており、他ノードの移動は全てこのルールに基づいて行なわれる。

4) テクスチャ・マッピング

変形を行なったWFMに対して、実際の人間の顔のテクスチャをはる。

6 合成システムのプロトタイプ

合成システムのプロトタイプは、SiliconGraphics社のグラフィックワークステーション上に実装され、

1) 表情の取得

2) WFMの変形

3) テクスチャ・マッピング

の部分を実現されており、現在は新たに視線と顔の向きの要素が追加されている。図9左は従来の無表情、右は視線を追加するために眼球を付加したものの無表情、図10はそれらの驚きの表情、図11左は顔の向きのみを、右は視線と顔の向きを追加した例である。

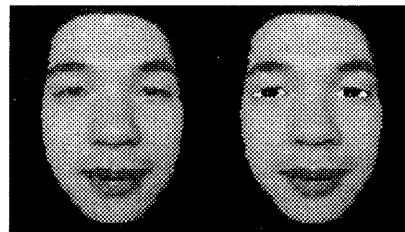


図9. 無表情

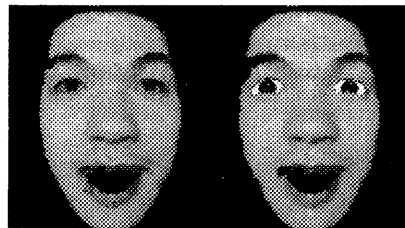


図10. 驚きの表情

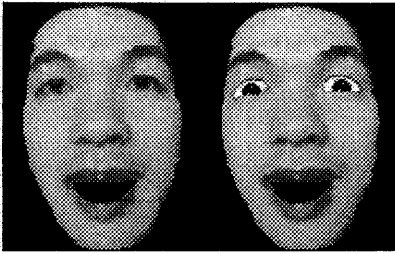


図 11. 顔の角度の追加

これら表情のみのものと、視線と顔の向きの要素を追加した表情による感情の認識についてアンケートによる評価を行なった。表 1 は表情のみによる感情の認識の結果、表 2 は視線と顔の向きを追加した表情による感情の認識の結果である。この結果によると喜びを除く表情では、正しく感情を認識できる割合が向上しており、視線と顔の向きを追加による効果が認められるが、喜びの表情では効果が得られなかったことが分かった。

| 表現された感情 | 認識された感情 | | | | | | |
|---------|---------|----|----|----|-----|----|-----|
| | 怒り | 嫌悪 | 恐れ | 喜び | 悲しみ | 驚き | その他 |
| 怒り | 4 | 1 | 0 | 0 | 0 | 4 | 1 |
| 嫌悪 | 1 | 4 | 0 | 0 | 0 | 3 | 2 |
| 恐れ | 0 | 0 | 1 | 2 | 1 | 2 | 4 |
| 喜び | 0 | 1 | 4 | 4 | 0 | 0 | 1 |
| 悲しみ | 0 | 1 | 0 | 2 | 2 | 0 | 5 |
| 驚き | 0 | 0 | 0 | 3 | 0 | 7 | 0 |

表 1. 表情変化のみの認識 (単位:人数)

| 表現された感情 | 認識された感情 | | | | | | |
|---------|---------|----|----|----|-----|----|-----|
| | 怒り | 嫌悪 | 恐れ | 喜び | 悲しみ | 驚き | その他 |
| 怒り | 6 | 3 | 0 | 0 | 1 | 0 | 0 |
| 嫌悪 | 2 | 7 | 0 | 0 | 0 | 0 | 1 |
| 恐れ | 0 | 2 | 2 | 0 | 0 | 3 | 3 |
| 喜び | 0 | 2 | 1 | 2 | 0 | 0 | 5 |
| 悲しみ | 0 | 1 | 0 | 1 | 4 | 0 | 4 |
| 驚き | 0 | 0 | 0 | 0 | 1 | 9 | 0 |

表 2. 視線と顔の向きの追加 (単位:人数)

7 まとめ

これまで感情通信システムの概要、分析システムにおける感情抽出の自動化へのアプローチ、合成システムにおける表情合成のための視線と顔の向きの要素の実装について述べた。

分析システムにおける特徴点追跡の方法を検討するために、今回は単純なテンプレートマッチングを用いたが、この方法では、周囲の明るさ等の要素による影響が非常に大きく、この点を解決できる方法の検討が必要であることがわかった。

合成システムでは、視線と顔の向きの要素を追加することである程度の感情の表現力の向上が確認された。しかし、効果の見られなかった感情もあり、

その原因のなどについて検討を進めている。また、合成された表情は不自然さを含んでおり、今後も改良の必要があると考えている。

参考文献

- [1] Shibata, Y., Kuraishi, H. and Sakamoto, H.: Feeling communication Feasibility using Facial Expression Analysis/Syntesis based on Individual Model, 情報処理学会論文誌, Vol.37, No.5, pp.920-930, May 1996.
- [2] W.Raffler, 本名信行 他訳: “ノンバーバル・コミュニケーション”, 大修館書店, 1981.
- [3] 伊藤, 富永; “表情の時間的変化を用いた個人識別手法”, 信学論 (D-II), Vol.J78-D-II, No.9, pp.1315-1324, Sep 1995.
- [4] 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, Aug 1994.
- [5] 森本, 李, 亀井, 井上: “個人差や撮影環境を考慮した濃淡画像からの口唇領域の抽出”, 信学論 (D-II), Vol.J79-D-II, No.1, pp.139-141, Jan 1996.
- [6] 山口, 福井: “分離度特徴を用いた顔画像解析”, 情報処理学会第 52 回全国大会論文誌, 1G-6, Mar 1996.
- [7] 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)
- [8] T.Inoue, K.Okada, Y.Matsusita, “Using gestural animations to convey the feelings in asynchronous network communication”, ICOIN-J005
- [9] 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)