

## ビジョンを備えた表情インタフェースとのインタラクション

内藤 剛人<sup>†</sup> 竹内 彰一<sup>‡</sup>

<sup>†</sup> 慶應義塾大学大学院 理工学研究科  
<sup>‡</sup> (株) ソニーコンピュータサイエンス研究所

表情を利用したインタフェースで、より自然なインタラクションをするにはユーザの認識が重要である。そこで、本報告では画像認識を表情インタフェースに導入したインタラクティブシステムについて報告する。まず、画像認識によりユーザの位置を求め、ユーザに対して自然な視線を向けることができる表情合成システムについて述べる。つぎに、ディスプレイの投影面が平面であることに起因するアイコンタクトの問題を解決するための2つのアプローチについて検討する。これらはともに、顔を回転させることで複数のユーザに対して選択的にアイコンタクトを行なう手法である。最後に、視線の応用例として2人のユーザがトランプゲームを行なうのをサポートするエージェントのシステムを示す。

## Interaction with Vision-based Synthesized Facial Display

Taketo Naito<sup>†</sup> Akikazu Takeuchi<sup>‡</sup>

<sup>†</sup> Department of Computer Science, Keio University  
3-14-1, Hiyoshi, Kohoku-ku, Yokohama, 223, Japan

<sup>‡</sup> Sony Computer Science Laboratory Inc.  
Takanawa Muse Building, 3-14-13 Higashi Gotanda  
Shinagawa-ku, 141, Japan

It is important for a synthesized facial display that it can really see the user. In this paper, we report about the interactive system that can recognize a position of the user, and synthesis facial displays with autonomous eye-direction. However, the user who does not stand just in front of the display cannot establish precise eye-contact selectively with the facial image, because the surface of the computer display is flat. We therefore make our two approaches to resolve that problem by rotating the face virtually or actually. Finally, we show an application of the interactive system that can communicate with two users.

## 1 はじめに

顔は人間にとってもっとも馴染み深いインタフェースである。そのため、顔をユーザとのインタラクションに利用するというのは極めて自然な発想である。加えて、近年は家庭用のコンピュータの三次元グラフィクスの処理能力も飛躍的に向上していることから、今後は多くのインタラクティブシステムに顔による表情インタフェースが用いられるようになることが予想される。

しかし、筆者らのこれまでの経験では、ただ画面の中のエージェントの顔が動くだけではユーザと自然なインタラクションを行なうのは困難であった。また、音声認識/合成を表情インタフェースと組み合わせたシステムでは、ユーザからの明示的な音声入力がない限りエージェントは反応することができなかった。そして、音声入力がないときは、果たしてユーザのことを本当に認識しているのかどうかをユーザにフィードバックすることも困難であった [8]。このような理由から、筆者らは表情インタフェースには画像認識による目が必要であると考え、顔が目を持つことで、ユーザに対して積極的に働きかけることができ、また視線によりエージェントの現在のアテンションがどこへ向けられているかを表すことも可能となる。

そこで本報告では、まず表情合成システムの概略について述べ、それから表情インタフェースに画像認識による目を持たせた試作システムについて報告する。また、このシステムを用いてユーザとのアイコンタクトを行なわせたところ、ディスプレイ投影面が平面であるためにどの方向から見ているユーザでもアイコンタクトができてしまうという問題が生じた。そこで、顔の向きを回転させることで現在のユーザを見ているかを明確に表現するという方法を2つのアプローチで行ない、検討を行なった。最後に、視線が効果的な役割を果たす例として、エージェントが2人のユーザがカードゲームを行なっているのを観察し、一緒にゲームに参加するという状況を作り、視線を備えた顔の影響について検討した。

## 2 表情合成システムの構成

表情合成システムはビジョン、アクション/リアクション生成、表情合成の3つのサブシステムで構成される(図1)。ビジョンサブシステムから入力される情報を元にアクション/リアクション生成サブシステムは、表情や視線、顔の向きを適時表情合成サブシステムに対して命令する。

これらのうち、ビジョンと表情合成サブシステムは本報告のすべての試作システムで共通なので、あらかじめここで述べることにする。

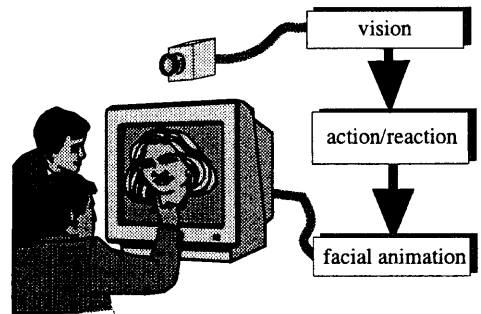


図1: システム構成

### 2.1 ビジョンサブシステム

視線を向けるために必要となる最低限の情報として、ユーザの位置が必要となる。そこで、ビジョンサブシステムでは単純な画像認識を行なっている。画像認識は次の4つの処理からなる(図2)。

- (1)解像度変換：処理を軽減するためにグラフィクスボードから送られる解像度  $646 \times 486$  ピクセルの画像を、 $32 \times 24$  に変換する。
- (2)差分計算：あらかじめ記録しておいた部屋の背景画像と入力画像の差分を計算することで、ユーザの存在する領域を求める。
- (3)ラベリング：個々の領域をラベルづけすることで複数のユーザを識別する。
- (4)重心計算：視線を向けるための点として、領域の重心を求める。

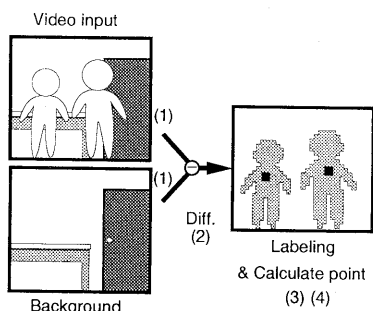


図 2: 画像処理の流れ

## 2.2 表情合成サブシステム

表情は、約 500 個の頂点からなる顔面ポリゴンモデルに対して、筋肉に基づく変形を加えることで合成される [7][12]。顔面モデルは、実際の人物の輪郭とテクスチャを取り込んだものを使用している (図 3)。また、筋肉は主要な 16 本の線状筋肉と 1 つの括約筋を、筋肉エディタを用いて顔面モデルに張り付けている [1][3]。

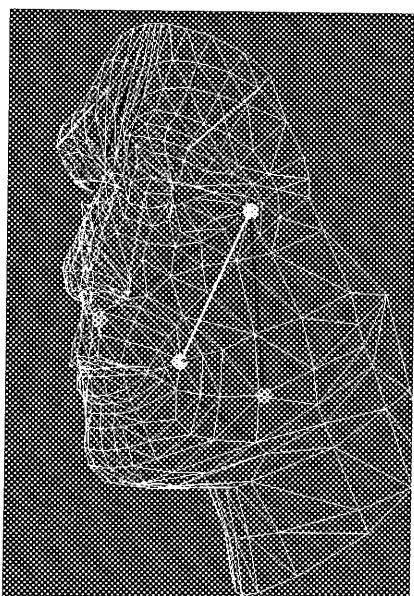


図 3: 顔面モデル

## 3 自然な視線を伴った表情合成システム

コンピュータエージェントに自然な視線を持たせる試みとして、ディスプレイに表示されたエージェントの顔が実際にユーザの顔を見て、動きを追跡するシステムの開発を行った。そこで、まずビジョンサブシステムから送られた情報をもとに視線の計算を行なうアルゴリズムについて述べ、本システムのためのアクション/リアクション生成サブシステムについて述べる。

### 3.1 ユーザへの視線の計算

視線は、眼球の中心点と瞳孔、そして観察する対象の 3 つが同一直線に乗るようにすることで形成される。このとき、両目の視線が交差する角度は輻輳と呼ばれ、輻輳角を作るための筋肉の緊張度は距離感の知覚に利用されている。一方、輻輳角は、相手が見ているのが遠くなのか近くなのかを知る際にも役立っている。エージェントが実際にユーザを見ているようにするためにはこの輻輳角が不可欠であるが、そのためにはユーザがディスプレイからどれだけ離れて座っているかを知る必要がある。ここでは 2.1 で述べたように単純な画像認識しか行っていないので、距離を 50[cm] と固定することにした。以下では、まずカメラ座標系におけるユーザの重心点までの角度計算について述べ、次にそれを眼球の回転角へ変換する計算について述べる。

#### (1) カメラ座標系における角度計算

カメラの焦点から入力画像上の任意の点  $(V_x, V_y)$  までの角度  $\theta_{cam}$  は次の式でそれぞれ求められる (図 4)。

$$\theta_{cam_x} = \tan^{-1} \frac{V_x}{f}, \quad \theta_{cam_y} = \tan^{-1} \frac{V_y}{f}$$

$f$ : レンズの焦点距離

#### (2) 眼球の回転角の計算

両目の  $x, y$  方向の角度  $\theta_x, \theta_y$  はそれぞれ以下のように求まる (図 5)。

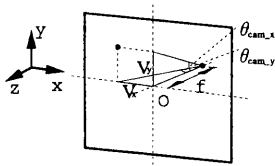


図 4: カメラ角度の算出

$$\theta_x = \tan^{-1} \frac{(p_z + d_{user}) \tan \theta_{cam_x} - c_x \pm \frac{d_{eye}}{2}}{d_{user} - c_z}$$

$$\theta_y = \tan^{-1} \frac{(p_z + d_{user}) \tan \theta_{cam_y} + p_y}{d_{user} - c_z}$$

$\vec{P}(p_x, p_y, p_z)$ : レンズ中央からディスプレイの投影面中央まで向かう変位ベクトル

$d_{user}$ : ユーザまでの距離 (50[cm] 固定)

$d_{eye}$ : 両目の間隔

$\vec{C}(c_x, c_y, c_z)$ : 顔の眉間 (両目の間) の座標

また、実世界の単位 [cm] をコンピュータ内部の正規座標系へ変換するためにディスプレイのドットピッチ (= 0.26[mm/dot]) とウィンドウ内のドットと正規座標との比率  $Ratio_x, Ratio_y$  を使用している。

### 3.2 アクション / リアクション生成サブシステム

本システムでは、基本的リアクションとして以下のものを行なえるようにした。このような少ないリアクションを規定するだけでも、エージェントに人間らしい印象を抱かせることができている。

#### (1) 視線および顔の向きによるユーザ追跡

ユーザが動くとき、その動きを視線および顔の向きで追跡する。このとき、目と顔を一緒に回転させると不自然になるので、目は常にユーザを追跡し、現在の顔の回転角  $\theta_{face}$  と目の回転角  $\theta_x$  の間に  $10^\circ$  前後の差がついた時に、顔を  $\theta_x - \theta_{face}$  だけ回転させるようにした。

#### (2) ユーザの認識のフィードバック

エージェントがユーザを認識していることを表情でフィードバックする。ユーザが視界にいるときは、絶えず口元に微笑を浮かべ、ユーザが視界からいなくなると、筋肉を全て弛緩させた無表情に戻す。

#### (3) 瞬き

実際の人間を参考に、約 10 秒おきに瞬きを行なうようにしている。また、(1) で顔を回転させる際にも 3 回に 1 回の割合で瞬きを行なうようにしている。前者は眼球の湿潤性を保つための瞬きを表現し、後者は外界の知覚の結果生じた瞬きを表現している [6]。

#### (4) 挨拶

対話の開始を知らせるために、視界に人が入ってきたら、その人の方向を向いて笑顔でおじぎをして挨拶を行なう。

### 4 アイコンタクトを行なう際の問題点

壁に貼られた人物ポスターなどを見ると、どの方向からも自分の方を見つめているような錯覚に陥ることがある [11]。こうした錯覚は、ディスプレイに表示された顔に対してでも同様に起きる。これは、顔面モデルがディスプレイの投影面に対して垂直に投影された二次元の画像として表示されているのが原因である。1 人のユーザを相手にする場合ならば、こうした錯覚は問題にならないが、複数のユーザを同時に相手する場合は現在誰とアイコンタクトをしているかが分かりづらくなり障害となる。

そこで、複数のユーザの中から選択的にアイコンタクトを行なえるような方法として、アイコンタクトをしたい相手に対して投影面を垂直にする、という方法を 2 つのアプローチで行ない、検討を行なった。

#### 4.1 画像変形による疑似的回転

ブラウン管ディスプレイのような重量のあるディスプレイの場合、直接回転させるのは

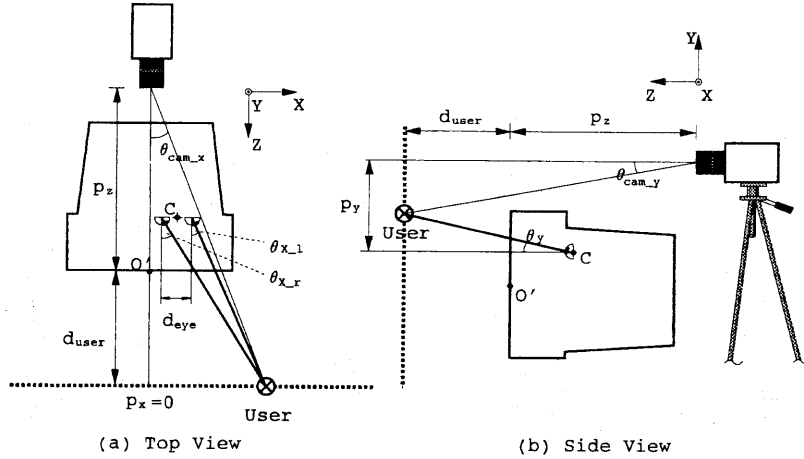


図 5: カメラ角度から目の回転角への変換

困難である。そこで、ディスプレイに投影する画像を変形し、疑似的に回転したよう表示してみることにした。

まずユーザが  $\theta_{view}$  の角度から見ている場合を考える。このとき、図 6 のようにディスプレイごと回転して、疑似投影面に対して顔画像が垂直に投影されたとする。これを、実際の投影面で切った断面が求める画像となる。ディスプレイ投影面の中央を原点  $O'$  とし、ディスプレイ内の三次元座標を  $(x, y, z)$  とすると、変形後の  $x$  座標は以下の式で表される。

$$x' = x - z \tan \theta_{view}$$

ところが、この変形だけでは  $\theta_{view}$  が大きくなると、ユーザから見て奥の部分ほど  $y$  軸方向の長さが短く見えてしまうことになる。そこで、頭部の中心  $H(H_x, H_y, H_z)$  を基準として奥の部分は大きく、手前のものは小さくなるように補正を行なう。変形式は以下の通りである。

$$y' = \frac{y \sqrt{\{x + (d_{user} - H_z) \tan \theta_{view}\}^2 + (d_{user} - z)^2}}{(d_{user} - H_z) \tan \theta_{view}}$$

これらの変形の結果が図 7 である。

しかし、実際に表示してみたところ、顔は多少横に歪んでもあまり違和感を感じることはないので、当初の目的としていた選択的な

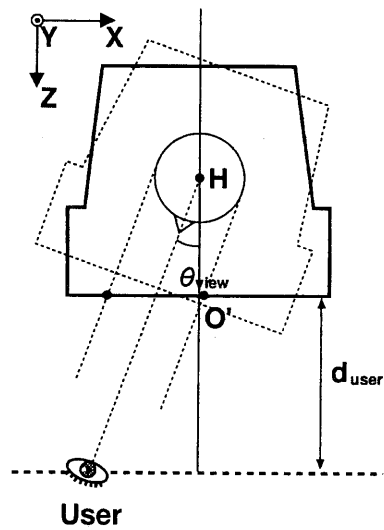


図 6: ディスプレイの疑似的な回転

アイコンタクトを正確に行なうには有効でないことが分かった。

## 4.2 モータによる回転

そこで、4.1とは対照的に、液晶ディスプレイを直接モータで駆動して、ユーザと対向させる方法をとってみた。図 8 で示されるよ



図 7: 歪ませた顔

うに、ステッピングモータ用のコントローラを WS のシリアルポートに接続し、表情合成サブシステムに顔の y 軸に関する回転角度を送らずに、代わりにコントローラへ送ることで顔の回転とした。

液晶ディスプレイは 4 インチのものを利用したので細かい表情を認識することが困難ではあったが、モータの回転音とともにユーザの方を向くという動作は、注意を引き付けることで大いに役に立ち、選択的なアイコンタクトにも適していることが分かった。

## 5 視線を利用したインタラクションの例

これまでのインタラクティブシステムの多くは 1 対 1 のインタラクションを想定していたが、筆者らはそれらを 1 対多、多対多に拡張したインタラクションに関心があり、社会的インタラクションの提案をしている [9]。社会的インタラクションでは、現在のアテンションがどこへ向けられているかを表す視線が大きな役割を果たしている。

そこで視線を利用したインタラクションの応用例として、2 人のユーザと 1 人のエージェントがカードゲームを行なう環境を作り、実験を行なった。

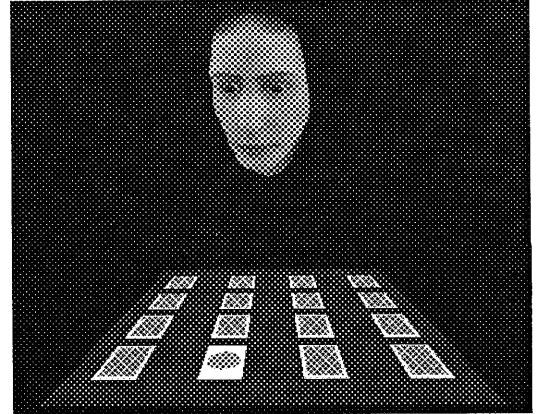


図 9: ゲーム画面

## 5.1 ゲームの設定

画面ではボード上にカードが置かれ、その上でエージェントが観察している (図 9)。カードは 4 種類のマークでそれぞれ 4 色ある。2 人のユーザはプレーヤとして、コンピュータの前にならんで座り、神経衰弱の対戦を行なう。エージェントは、直接ゲームに参加するのではなく、見物客という立場をとる。しかし、エージェントはただ観戦するだけでなく、一方のプレーヤの味方につき、視線で当たりのカードを示唆することで加勢する。

ゲームを開始すると、プレーヤはカードを 2 枚ずつめくり、同じマークのカードならばボード上から消えて得点が加算される。通常の神経衰弱とは異なり、カードを正解しても順番は次のプレーヤに移る。これをカードがなくなるまで繰り返す。

エージェントはゲームの進行中に、ゲームの進行に応じたアクション (表 1) や、プレーヤの行動に応じたリアクション (表 2) を表情と顔の向きおよび視線で行なう。

## 5.2 被験者による実験

4 組のペアに実際にこのシステムを利用してもらった。その結果、興味深い結果がいくつかえられた。

まず、ゲーム中におけるプレーヤの視線の

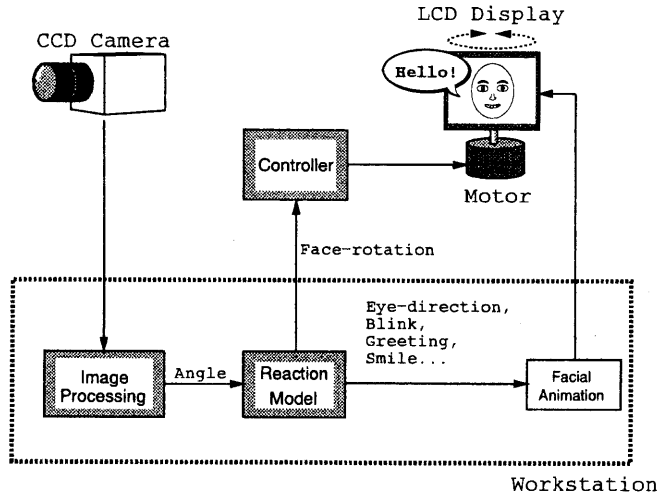


図 8: モータ回転によるシステム

表 1: ゲームの進行に応じたアクション

| Game Actions        |
|---------------------|
| 現在の順番のプレイヤーを見る。     |
| 味方にペアになるカードを教える。    |
| 正しいカードにマウスカーソルを乗せた。 |
| 正しいカードを開いた。         |
| 間違ったカードを開いた。        |
| 味方が得点した。            |
| 敵が得点した。             |
| 味方が勝利した。            |
| 引き分け。               |
| 敵が勝利した。             |

動きを観察した。すると、エージェントが味方についてプレイヤーの方が、エージェントを長い時間見ていたことが分かった。また、エージェントとアイコンタクトが成功する割合も高いことが分かった。これは、エージェントがプレイヤーの扱いに差をつけたことが原因である。このことは、アイコンタクトが重要な意味を持つような状況ではユーザへのリアクションを慎重に設計しなければならないことを示唆している。

表 2: プレーヤの行動に応じたリアクション

| Situated Reactions   |
|----------------------|
| プレイヤーの動きを追う。         |
| マウスカーソルの触れているカードを見る。 |
| プレイヤーに操作を促す。         |
| 大きく動いたプレイヤーの方を見る。    |
| 視界からプレイヤーが消えた時。      |

また、相手の順番のときにプレイヤーが大きく動くエージェントはそちらの方を向くリアクションをとるようにしたが(表2の4行目)、その際に、プレイヤーが無意識にエージェントの方を見てアイコンタクトをとってしまう現象が多く見られた。この原因として顔の持つ特性が考えられる。人は顔を自分の方に向けられると、たとえそれがエージェントの顔であってもついそちらを見てしまうのである。このように、擬人的なインタフェースを用いると人間同士のインタラクションのときの反応を喚起できることが分かった[5][10]。

## 6 おわりに

表情合成システムにカメラを接続し、目としての働きを持たせることで、ユーザの動きを認識し、自然な視線を伴った表情で反応するリアクションシステムの試作を行なった。

さらに、複数のユーザとのインタラクションを考慮した場合、通常のディスプレイでは正確なアイコンタクトを行なうのは困難である。そこで、顔画像を歪ませることによる疑似的な顔の回転方法と、モータに据え付けた液晶ディスプレイを回転させる方法の二つを試みた。

最後に、視線が大きな役割を果たすインタラクションの例として、2人のユーザがカードゲームを行なっている環境にゲームの進行を補助する役割のエージェントが存在するようなシステムを試作して、実験を行なった。その結果、ユーザの立場の違いによりアイコンタクトの頻度にも差が出る事が分かった。また、顔という擬人的なインタフェースを用いると人間同士のインタラクションのときのような反応を無意識に引き出すことがあることも分かった。

今後は、複数のエージェントと複数のユーザが混在するような多対多のインタラクションへの拡張を予定している。

## 参考文献

- [1] Ekman, P. and Friesen, W. V. (工藤 力 邦訳): 表情分析入門, 誠信書房, 1987.
- [2] Morii, K., Satoh, T., Tetsutani, N. and Kishino, F.: A Technique of Eye-Animation Generated by CG, and Evaluation of Eye-Contact Using Eye-Animation, *Proc.SPIE Conf. on Visual Comm. and Image Processing '92*, pp.1350-1357, 1992.
- [3] 内藤 剛人, 竹内 彰一, 所 真理雄: 筋肉エディタによる表情アニメーションの向上, 情報処理学会 グラフィクスとCAD シンポジウム, pp.69-78, 1993.
- [4] 内藤剛人, 竹内彰一, 所真理雄: 視線を伴った表情合成システム, インタラクティブシステムとソフトウェア I: 日本ソフトウェア科学会 WISS '93(竹内彰一(編)), 近代科学社, pp.201-208, 1994.
- [5] Nass, C., Steuer, J. and Tauber, E. R.: Computers are Social Actors, *Proc. CHI'94 Human*

*Factors in Computing Systems*, ACM Press, pp.72-78, 1994.

- [6] 多田英興, 山田富美雄, 福田恭介: まばたきの心理学, 北大路書房, 1991.
- [7] Takeuchi, A. and Franks, S.: A Rapid Face Construction Lab, *SCSL-TR-92-010*, Sony Computer Science Laboratory, Inc. Tokyo, 1992.
- [8] Takeuchi, A. and Nagao, K.: Communicative Facial Displays as a New Conversational Modality, *Proc. INTERCHI'93*, ACM Press, pp.187-193, 1993.
- [9] Takeuchi, A. and Naito, T.: Situated Facial Displays: Towards Social Interaction, *Proc. CHI'95 Human Factors in Computing Systems*, ACM Press, pp.450-455, 1995.
- [10] Walker, J., Sproull, L. and Subramani, R.: Using a Human Face in an Interface, *Proc. CHI'94 Human Factors in Computing Systems*, ACM Press, pp.85-91, 1994.
- [11] Warahha, H.: 頭を動かしても外界はなぜ静止して見えるか, 日経サイエンス, Vol.7, pp.72-79, 1985.
- [12] Waters, K.: A Muscle Model For Animating Three-Dimensional Facial Expression, *Computer Graphics*, Vol.22, No.4, pp.17-24, 1987.
- [13] 谷内田 正彦: ロボットビジョン, 昭晃堂, 1990.