

人間とコンピュータとの表情コミュニケーションシステムに関する研究

(知能情報システム学) 辻光裕

1. 序言

人と人とのコミュニケーションにおいて、顔の表情が果たす役割は大きいと考えられる。ゲームや映画においては、登場するバーチャルキャラクターの表情が、ゲームプレイヤーや視聴者の作品に対する評価を左右する。近年、表情やコミュニケーションに関する研究は盛んに行われており、その分野は心理学、認知科学、ロボット工学、CG など広範囲に及んでいる。情報処理振興事業協会 (IPA) では表情合成プログラム (FaceTool [1]) をフリーウェアとして公開しており、このソフトウェアの顔認識・合成についての報告もある。

一方、人は他者とのコミュニケーションでは、相手の表情に対して相互に同調的な表情を表出することがある。この現象は同調的反応と呼ばれ[2]、多くの研究報告がある。人とコンピュータとの表情コミュニケーションにおける同調的反応の、後続する表情に与える影響についての研究[2]において、人間の方から一方的に表情 (表情 A) を表出し、表情 A に対して仮想他者 (バーチャルキャラクター) が表情 (表情 B) を表出する応答をした場合の、表情 B に対する人間の表情の変化が報告されている。しかし、この研究では、頭部に固定したカメラから被験者の画像を取り込んでいるなど、環境的制約を与えている。そこで、本研究では、被験者に何ら環境的制約を与えずに表情を認識し、かつ、仮想他者に表情を表出させ、人間とコンピュータとのコミュニケーションの実現を図るシステムを開発し、その性能評価を行った。

2. 処理概要

本システムは表情学習部、表情認識、表情表出部から成り、各構成における処理概要を以下に示す。また、画像入力にはディスプレイ上に設置したソニー製のデジタルビデオカメラ (DCR-TRV20) を用い、バーチャルキャラクターの CG 作成には IPA の表情合成プログラム (FaceTool [1]) を用いる。なお、被験者については脱帽・眼鏡取り外し等の制約はなく、蛍光灯下の室内でシステムを稼働させている。

【表情学習部】

- (1) ビデオカメラから取り込んだ、瞬きによる目の開閉の画像を用いて、目領域テンプレート及び平均無表情顔画像を作成する。
- (2) 表情 (笑顔、無表情) を表出した顔画像 (表情画像) に対して、テンプレートマッチングを施し、表情特徴を表す顔部品を抽出する。
- (3) 平均無表情顔画像と表情画像との差分に対して 2 次元離散的コサイン変換 (2D-DCT) を施し、各周波数帯域において DCT 係数の絶対値の各顔部品領域における平均値を表情特徴量として求め、学習データとする。

【表情認識部】

- (4) (1) ~ (3) と同様の処理を行い、認識用の画像から得られた表情特徴量と学習データとの特徴ベクトル空間での距離から表情を認識する。

【表情表出部】

- (5) (4) の認識処理で得られる表情の強度を用いて好印象合成顔[3]を変化させ、表出された表情に対する被験者の表情反応を記録する。なお、仮想他者の表情パターンは【A、仮想他者が独立で笑う】【B、人間の表情に対応した強度で笑う】【C、A と B の混合】の 3 種類とした。

3. 表情学習部及び表情認識部

目領域の検出には、3 回の瞬きによる両目の開閉画像を用いる。まず、目を開いた BMP 画像と目を閉じた BMP 画像との輝度の差分画像を作成する (図 1)。



図 1 目を開けた画像 (左) と目を閉じた画像 (中) との輝度の差分画像 (右)

作成した差分画像に対して二値化処理を施し、顔の中心付近を抽出し、ピクセルの膨張処理と収縮処理を施して微小なピクセルを除去し、両目の候補となるピクセルのクラスター群を作成する。各クラスターに対して、ラベリング処理を行い、画像内の各左右でクラスターの重心を求め、左右のクラスター群の重心を仮の両目候補とし、この点から最近傍の各左右のクラスターを両目の候補とする。

両目候補を結んだ直線と水平方向との角度 θ を求め、水平線と直線との交点を両目の中心とする。両目の中心をアフィン変換で平行移動し、画像の中心と一致させた後、アフィン変換で画像を角度 θ だけ回転させ、顔の傾きの補正を行った後、両目の中心間距離を一定とするアフィン変換 (サイズ規格化) を画像全体に施し、しかる後、図 2 のような目領域テンプレートを作成する。

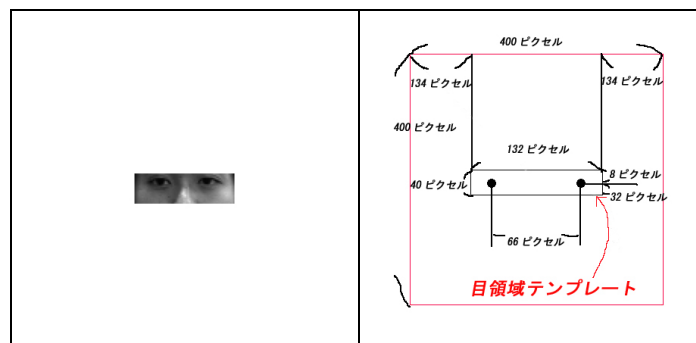
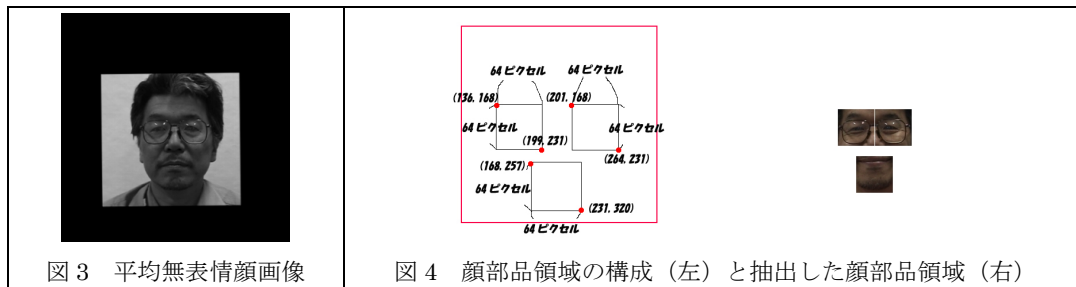


図 2 目領域テンプレートとその構成

目領域テンプレート作成の際の【画像中心への平行移動距離】、【回転角度】、【サイズ規格化の際の拡大・縮小率】を用いて、瞬きの開眼画像 3 枚のそれぞれに対して同様のアフィン変換 (以下では「規格化」と表記) を行う。更に、目領域テンプレートを用いて、3 枚の開眼画像とのテンプレートマッチングを施す。この処理の際のアフィン変換の条件決定にはアニーリング処理を用いる。そして、得られたアフィン変換後の 3 枚の開眼画像の平均化を行うことにより、平均

無表情顔画像を作成する（図 3）。

平均無表情顔画像の作成と同様に、ビデオカメラから入力された表情画像に対して画像の規格化及び上記テンプレートマッチングを施し、解析対象となる顔部品の各矩型領域（【右目・右眉領域】、【左目・左眉領域】、【口・アゴ領域】）の抽出を行う（図 4）。



各顔部品領域で 8×8 画素サイズの各領域に対して各々 2D-DCT を施し、各周波数帯域において 2D-DCT 係数の絶対値の各顔部品領域における平均値を 1 つの表情特徴量（DCT 特徴量）とし、合計 45 個の DCT 特徴量を得る。これを特徴ベクトルとして学習データを作成し、特徴ベクトル空間での最小距離識別法による表情認識を行う。

4. 表情表出部

本システムでは、バーチャルキャラクターの表情合成に、IPA の【FaceExp】とアンケートを用いた表情合成[3]の手法を用いており、笑顔の表情に好印象顔合成ベクトルによって合成される、誰にでも好印象を与える可能性の高い表情の【好印象合成顔[3]】を採用している。表情認識処理の際に特徴ベクトル空間上で認識用の表情画像と最も近い表情パターンを持つ学習画像（以下、「最近接類似学習画像」と表記）とそれ以外の表情カテゴリーで認識用の表情画像に最も近い表情パターンを持つ学習画像（以下、「最近接相違学習画像」と表記）をあわせて求め、この 2 つの特徴ベクトルを用いて、以下のようにバーチャルキャラクターの表情表出の強度を定義する。

特徴ベクトル空間上で最近接類似学習画像の点と最近接相違学習画像の点を結ぶ直線に、認識用の表情画像の点から垂線を下ろし、垂線と上記の直線との交点と最近接類似学習画像の点及び最近接相違学習画像の点との距離の比を計算する（図 5）。この内分比 (a, b) を用いて、バーチャルキャラクターの表情表出の強度 $\left(\frac{b}{a+b} \times 100\right)$ に変化をつける。

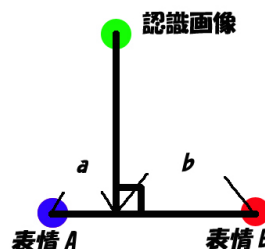


図 5 特徴ベクトルの距離の比

5. 実験結果及び考察

5.1 表情認識実験

20 代男子 8 人と 40 代男性 2 人にビデオカメラの前に座ってもらい、各々から 3 回の瞬きによる目開閉の画像 3 組（開眼画像 3 枚、閉眼画像 3 枚）と【無表情】と【笑顔】の表情画像をそれぞれ 10 パターンずつ採取し、学習データとした。採取された画像はそれぞれ合計で、瞬き画像 60 枚（開眼画像 30 枚、閉眼画像 30 枚）、無表情顔画像 100 枚、笑顔画像 100 枚となった。

続いて、学習データ採取時と同様に、上記 10 人に瞬きを 3 回、【無表情】と【笑顔】を 10 パ

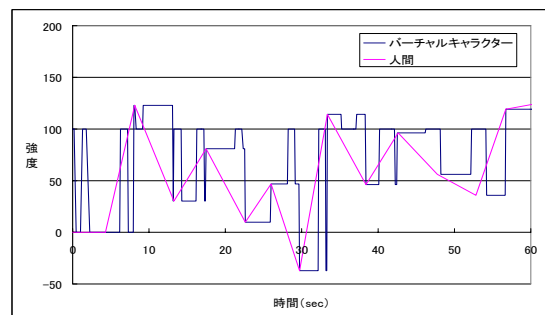
ターンずつ表出してもらい、リアルタイムに表情認識を行った。

表情認識率は無表情 91%、笑顔 90%、平均 90.5%であった。誤認識が生じたのは、学習データ作成時や認識データ採取時に三次元的な顔の回転（傾きや首振りのような顔の動き）が生じ、平均無表情顔画像との差分が大きくなったためと考えられる。

5.2 表情応答実験

年齢 18～22 歳の男性 8 人、女性 6 人と 40 代男性 1 人にビデオカメラを置いたテレビの前に座ってもらい、表情認識実験と同様に【無表情】と【笑顔】をそれぞれ 10 パターンずつ学習データとして採取し、被験者に「CG キャラクターがテレビに現れて表情を変えますので、キャラクターを見ながら自由な表情で応じて下さい」と教示した上で、テレビにバーチャルキャラクターを表示し、被験者の表情変化を記録した。バーチャルキャラクターの表情パターンは、【A: (ランダムなタイミングで) バーチャルキャラクターが独立に笑う】、【B: 人間の表情強度で笑う】、【C: A と B の混合】の 3 種類とし、それぞれ 1 分ずつ実験した。

図 6 に上記 C パターンの実験結果例を示す。縦軸の値で 50 以上が【笑顔】、50 未満が【無表情】を表しており、バーチャルキャラクターが表出した笑顔の表情に対して、被験者はそれに誘起され笑顔となる現象が観察された。図からもバーチャルキャラクターの笑顔に対して同調的に被験者に笑顔が現れることが分かる。



また、A, B パターンについても、同様の現象が観察された。実験後に行ったアンケートの「CG キャラクターが自分の表情に反応した事について」の項目には、「嬉しかった」、「笑えた」、「癒された」などの回答があった。

なお、本実験全体での顔部品領域抽出不良率は平均 21.38%であった。顔部品領域抽出不良の場合には、表情は全て【笑顔】と認識されていた。また、【無表情】と認識された表情のうち、4.76%が笑顔以外の表情であり、【笑顔】と認識された表情のうち、18.53%が笑顔以外の表情であった。

6. 結言

実験後のアンケートから、本システムには認識系の更なる高速化が必要であることが分かった。また、より円滑な表情コミュニケーションのため、今後は笑顔以外の表情も認識及び表出させる予定である。

参考文献

- [1] <http://www.tokyo.image-lab.or.jp/aa/ipa/>
- [2] 市川寛子, N.P.チャンドラシリ, 原島博, 牧野順四郎, “相手の同調的反応が観察者表情に及ぼす影響”, 信学技法 PRMU2002-210 (2003-02).
- [3] 辻光裕, 吉富康成, 谷尻豊寿, “アンケートを利用した表情合成とその評価”, 日本顔学会誌, 第 3 巻, 第 1 号, pp. 111 - 117 (2003).