

顔表情の知的符号化方式の検討 - リアルタイム通信の可能性 -

Examination intelligent facial coding system

-Possibility of real-time communication-

伊豆川悠介 糸井清晃 小林幸雄

Yusuke Izukawa Kiyooki Itoi Yukio Kobayashi

千葉工業大学

Chiba Institute of Technology

1 まえがき

顔には、その人物の個性を表現するディスプレイ、時々刻々の心理状態を反映させたインジケータ、そして伝達すべき意図を表現するメッセージという働きがある[1]。人は日々のコミュニケーションの中で時々刻々と変化する喜びや怒りといった感情を無意識のうちに、あるいは表現するために顔の表情筋を動かす。これにより平常顔から変化した表情を相手に示し、自分の心理状態を伝えることができるのである。遠隔地とこのようなコミュニケーションをとる場合テレビ電話が用いられる。しかし、カメラより取り込んだ 320×240 のカラー画像の送受信を行う場合、 $6 \cdot 6 (= 320 \times 240 \times 3 \times 30)$ (Mb/sec) の伝送速度が必要とされる。テレビ電話で必要な映像情報とは、取り込んだ映像そのものではなく、通話者の表情であると考えられる。

そこで本研究では、表情によって感情を伝える顔の役割に着目し、表情パラメータのみを伝送する知的符号化方式[2]の検討、及びリアルタイム通信の可能性について検討を行う。

2 原理

本研究はカメラより取り込んだ映像において画像データそのものを符号化し送受信するものではない。送信するデータは認識部による表情の認識結果である(図1)。受信側では送信されたデータをもとに表情生成を行う(図2)。送信側および受信側の詳細は

2.1 および 2.2 にて後述する。

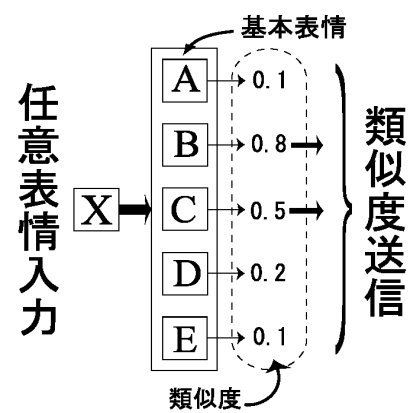


図1 認識部概要

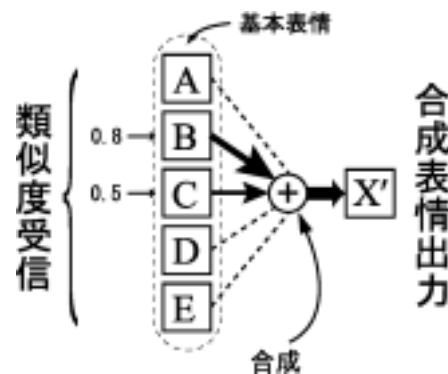


図2 合成部概要

2.1 送信側(認識部)

送信側では基本表情画像を学習したニューラルネットで任意表情画像の認識結果の出力を行う。基本表情とは合成部において認識表情を生成する為に必要な顔表情のことである。本研究では図3に示すような人

が対話するときに必要な 5 つの表情を用いた。



(左から平常・笑う・泣く・怒る・驚く)。

図 3 基本表情画像

ニューラルネットの入力層には表情をモノクロ化、明度補正およびモザイク化した際の各ブロックの値を 0~1 に変換して入力する。このとき各器官に着目したモザイク処理(目周辺：各 100，口周辺：100)を行うため、入力ユニット数は 300 とした。中間層は 64 とした。出力ユニット数は 5 表情画像を 2 枚ずつ用意し、それぞれ区別するために 10 とした。このニューラルネットに基本表情をバックプロパゲーション学習させた。これに任意表情を認識させ、そのときの出力値を類似度として送信する。

2.2 受信側(合成部)

あらかじめ基本表情画像を参考に特徴点で表現した画像を用意する(図 4 参照)。合成部では受信した類似度より、表情を合成し出力する。その類似度が上位 2 位までの表示用似顔絵画像を以下の式を用いて座標計算を行う。

$$P_0(x, y) = \{A \times P_1(x, y) + B \times P_2(x, y)\} / (A + B)$$

A: 上位 1 位の類似度

B: 上位 2 位の類似度

P_0 : 合成後の座標

P_1 : 上位 1 位の似顔絵座標

P_2 : 上位 2 位の似顔絵座標



図 4 表示用似顔絵

画像の生成には合成部によって求めた座標

の計算結果より、各点に対してスプライン補間を行う。

3. 実験と結果

3.1 実験

従来の手法では 1 つの表情に対して 1 枚の基本表情画像を用意していた。モノクロ化 明度補正 12×12 モザイクの画像処理を行っていたが、特定の表情のときにしか正しい結果が得られなかった。そこで幅広く表情認識できるように 1 つの表情に対して 2 枚の画像を用意した。

3.2 結果

図 5 は第 3 章で述べた手法での認識結果である。認識用顔画像は基本表情画像の撮影とは別の日に各表情 2 枚ずつ撮影したものをを用いた。表の数値は 2 回の認識結果の平均値である。

	平常	笑う	泣く	怒る	驚く
平常	0.995994	0.001495	0.000344	0.001977	0.000326
笑う	0.005001	0.881284	0.002288	0.010443	0.000028
泣く	0.004047	0.000042	0.963628	0.003379	0.001078
怒る	0.000388	0.000232	0.003828	0.874789	0.001656
驚く	0.000665	0.000006	0.001875	0.001418	0.986424

図 5 表情認識結果

4. あとがき

速い表情認識処理(0.2(sec))と高い認識率が得られたため、リアルタイム通信を行うための有効的な手段であるといえる。今後の課題は受信側における画像生成過程の改良である。

参考文献

- [1] 赤松 茂 “人間とコンピュータによる顔表情の認識” 電子情報通信学会誌 Vol.85, No.9, pp.680-685, Sep, 2002
- [2] 御園生 靖史 “NN とモーフィングを用いた顔表情知的符号化の研究” 千葉工業大学, 2000