

感情マップによるリアルタイム表情合成の試み

宮崎 光二*

An Attempt of Real-Time Facial Expression Synthesis Using Emotion Map

Koji MIYAZAKI*

ABSTRACT

In recent years, technological development in the IT field related to the face has progressed. In the past, it was only possible to detect the face part from the image, and it was not easy to implement practical functions. It is becoming possible to use it. In this paper, we will create a mechanism that can operate multidimensional parameters of face information in a dimension that humans can intuitively recognize, and try to develop a UI and facial expression synthesis system that creates facial expressions with CG in real time. The main parts that make up the facial expression are the eyes, eyebrows, nose, and mouth, and the facial expression can be expressed by moving or deforming these. However, it is very difficult to specify many parameters such as the position / tilt, size, and shape of these facial parts at the same time and control them appropriately. In this paper, we reduce the number of dimensions of many parameters and realize a mechanism that can be controlled in an emotional parameter space that can be expressed in two or three dimensions by using a neural network.

キーワード：顔表情，ニューラルネットワーク，感情マップ，3DCG

Keywords: Facial Expression, Neural Network, Emotion Map, 3DCG

1. まえがき

近年，顔に関する IT 分野の技術開発が進んでいる。以前は画像から顔部分を検出する程度であり，実用的な機能の実装は簡単ではなかったが，カメラおよび形状認識系センサーにおける各分野での性能向上が進み，顔进行处理するシステムの実用的な利用が可能となってきた。顔に関する情報を処理するライブラリやフレームワークも多く存在し，さまざまなシステムへの組み込みが容易になっている。顔認識技術利用の身近なところでは，Apple 社の iPhone でロック解除に顔認証システム FaceID を採用しており，これは赤外線照射してその反射状況を特殊なカメラで取得することで顔の形状を認識している。Apple はさらに簡

単に 3DCG で顔表情を制御できる「ミー文字」を開発し，チャットアプリやさまざまなコミュニケーションツールにおいて利用しやすいしくみを提供している。「ミー文字」は自分の顔を撮影しているカメラによって顔の形状の情報を取得し，自分の表情をリアルタイムに「ミー文字」のキャラクターに反映させることができる。「ミー文字」に利用される顔形状情報の取得は，赤外線カメラおよび光学カメラから得られた映像を高性能なニューラルエンジン搭載チップで統合・処理することにより実現される。複数の高性能カメラの利用と高度な情報処理により，自分の顔表情をダイレクトに CG キャラクターに反映させる機能を可能にしている。顔の表情合成に関する研究は数多くされおり，「ミー文字」のように実際の顔形状を忠

実に反映させるものや、似顔絵の要素を取り入れてより表情豊かに表現することを目的としたもの[1]、また感情モデルに連動させたもの[2]など、多くのアプローチが存在する。顔表情を表現するには多くのパラメータが必要であり、表情を形成するポイントとなる顔の部位だけでも目、眉、口があり、それぞれにおいて複雑な変形処理を行う必要がある。前述の「ミー文字」のように高性能なカメラと高性能チップを用いて大量の情報を高速処理すればリアルタイムかつリアルに表情を合成することが可能であるが、ハードリソースが乏しい場合や同時に多数の顔表情を合成する場合は処理が困難になる。ソフトウェア的な処理の負荷が小さい合成手法は教育支援システム[3]やネットワークを介したコミュニケーションなどにおいて応用範囲が広く、利用価値があると考えられる。本稿では顔情報の多次元パラメータを人間が直観的に認識できる次元で操作可能なしくみを作成し、リアルタイムにCGで顔表情を創出するUIおよび表情合成システムの開発を試みる。顔の表情を構成する主な部位は、目、眉、鼻、口であり、これらを移動・変形させることにより表情を表すことができる。しかし、これらの顔部位の位置・傾き、大きさ、形状など、多くのパラメータを同時に指定して、適切に制御することは非常に困難である。本研究では、多くのパラメータの次元数を減少させて、2次元または3次元で表現可能な感情パラメータ空間で制御できるようなしくみを、ニューラルネットワークを利用することで実現する。

2. システム構成

2. 1 概要

顔部位の形状パラメータと、2次元の感情パラメータでニューラルネットワークの学習を行い、その結果を用いて顔表情を2次元座標空間をポイントするだけでその感情パラメータの表情を3DCGで描画するシステムを開発した。図1はそのイメージ図であり、左側の簡単な感情マップ上である。感情マップでは中

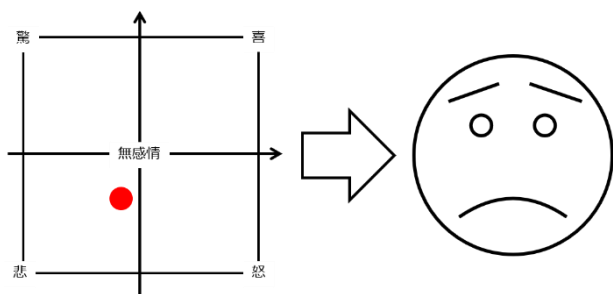


図1 感情マップと表情作成

Fig.1 Emotion Map and Facial Expression Synthesis.

心を原点として、感情がない（無表情）状態を意味する。簡易的に4象限のそれぞれに基本的な4つの感情を割り当て、感情マップの任意の場所でポイントすると、その位置の(x, y)座標の値に対応する表情をリアルタイムに合成・描画する。

2. 2 顔の描画

顔は3DCGで、単純なSphereとCylinderの組み合わせで顔のベースと、眉、目、口を構成する。ポリゴンの変形や補完処理などは行わず、描画や計算の手順をできる限り単純にして、処理の負荷を小さくする。図2は顔の描画例である。

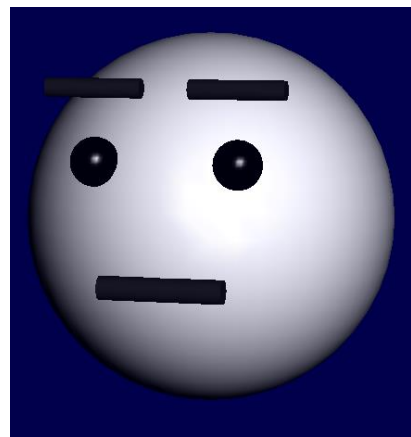


図2 基本的な顔の描画例

Fig.2 Example of Face CG.

顔の表情は、眉の位置と角度、目の大きさと縦横の長さ、口の位置と形で表現する。口は真ん中から左右に2個のCylinderを配置し、「へ」の字のように折れ曲がるしくみになっている。また、表情を変化させるパラメータは、以下の6個とした。

- (1) 目のヨコのサイズ
- (2) 目のタテのサイズ
- (3) 眉のタテの位置（変化量）
- (4) 眉の角度
- (5) 口のタテの位置（変化量）
- (6) 口の角度

無表情（無感情）の時に上記(1)～(6)の各パラメータは初期値となる。各パラメータの値は連続値であり、値が小さく変化すれば表情は少し変わり、値が大きく変化すれば表情も大きく変わる。各パラメータを適切に設定すれば、意図する表情を作成することができるが、6個のパラメータに対して適切な値を手作業で設定することは難しい。そこで、本システムでは感情マップの2次元情報を入力値として、6個のパラメータ値を出力値とするニューラルネットワークを構成することにより、顔表情をリアルタイムに合成・創出す

るシステムを開発する。

2. 3 感情と表情

人の顔表情は、感情の変化に伴い同様に变化する。感情が無い(ニュートラルの)場合、表情は無表情となる。怒っているときは、表情も怒り顔となり、悲しいときは悲しそうな顔となる。感情の状態に連動して表情が変わるしくみは自然なことであり、表情をコントロールする方法として直観的に理解しやすい。感情を表現するモデルはいくつかあり、その中で代表的なものに「ラッセル円環モデル」(図3)がある[4]。これはラッセルが提唱した快-不快、覚醒-眠気の2次元空間上に情動を定義できるとしたモデルである。2次元空間上に2つの軸の程度を考慮して情動ワードを円状に配置したものである。ラッセル円環モデルは心理学や情動研究において多くの研究で使用されているが、28個の情動ワードを並べて感情を表現すること

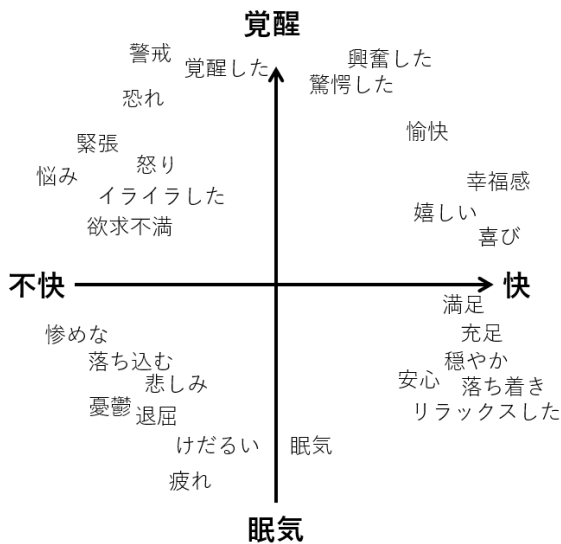


図3 ラッセルの円環モデル

Fig.3 Russell's Circumplex Model.

から、本システムでの利用には適さない。

感情・表情研究の第一人者である Ekman らによれば、人間の持つ基本表情は大きく分けて「幸福、驚き、怒り、恐怖、嫌悪、悲しみ」の6種類に分類できると言われている[5]。この6種類のなかで、恐怖と嫌悪は表情の特徴が似ているところが多く、また眉、目、口などの部位の変位が小さいため、表情からその感情を推し量ることが難しい。本システムは感情の要素を2次元空間で表し、感情を直観的にポイントおよび変更することにより表情の状態をコントロールする。感情マップでは、4種類の感情と無感情の5種類の感情

を基準とし、4種類の感情は表情の違いが大きいと思われる幸福、驚き、怒り、悲しみを採用する。

2. 4 ニューラルネットワーク

感情マップ上でポイントされた位置の(x, y)座標から6個のパラメータへの変換は、ニューラルネットワークを用いる。ニューラルネットワークは3層構造(図5)であり、ユニット数は入力層2個、中間層7個、出力層6個である。教師データについて、まず幸福、驚き、怒り、悲しみ、無感情の5個の表情を手作業で作成し(図4)、それぞれの6個のパラメータを決定し、それらを出力層ユニットの教師データとする。次に、感情マップ上で幸福、驚き、怒り、悲しみ、無感情を(1, 1), (-1, 1), (1, -1), (-1, -1), (0, 0)の座標に配置するものとして、これらの値を入力層ユニットの教師データとする。教師データは正規化した値を用いて、ニューラルネットワークの学習はバックプロパゲーションで行った、

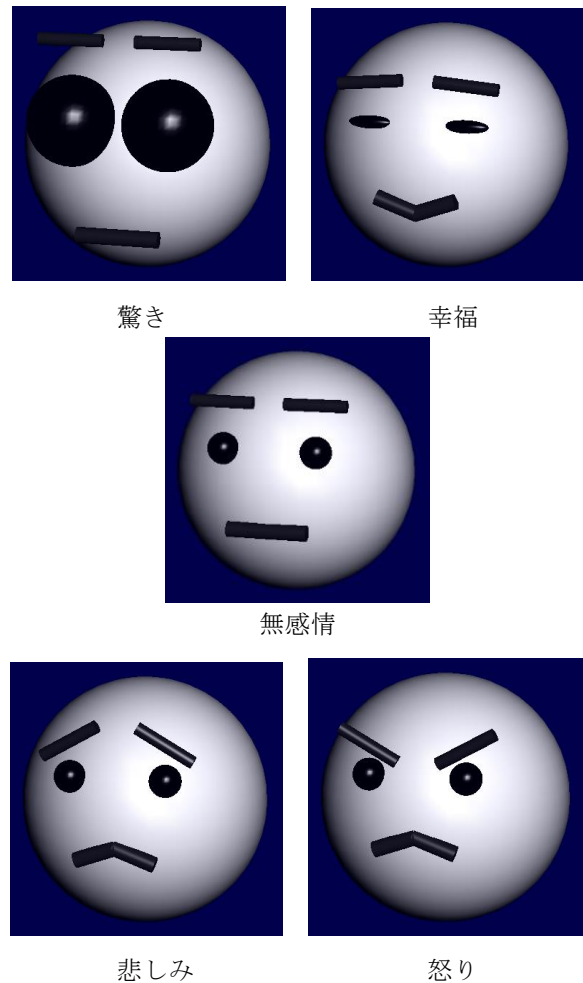


図4 教師データとなる顔表情

Fig.4 Teaching data of face expression.

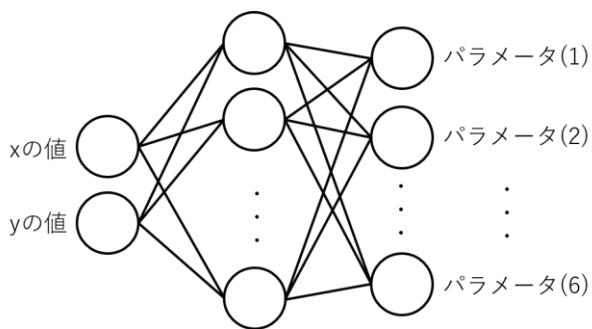


図 5 ニューラルネットワーク

Fig.5 Neural Network.

3. 実行例

本システムの実行例を示す。プログラムを実行すると、感情マップと顔表情を描画するウィンドウが表示される。感情マップでマウスをポイントすると、その位置の x, y 座標がニューラルネットワークの入力値となり、リアルタイムに出力値である 6 個のパラメータを計算、その値を使って顔表情を 3DCG で描画する。図 6 は無感情（無表情）であり、感情マップ上で

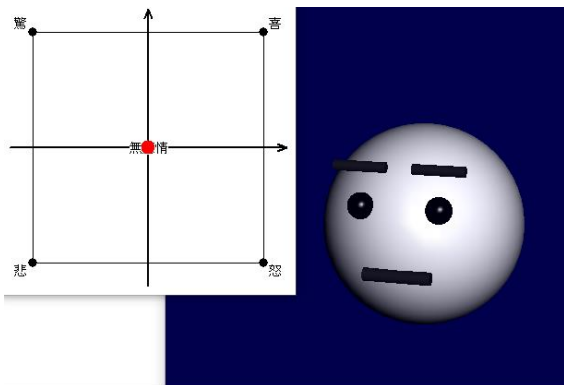


図 6 実行結果(無感情)

Fig.6 Execution result.

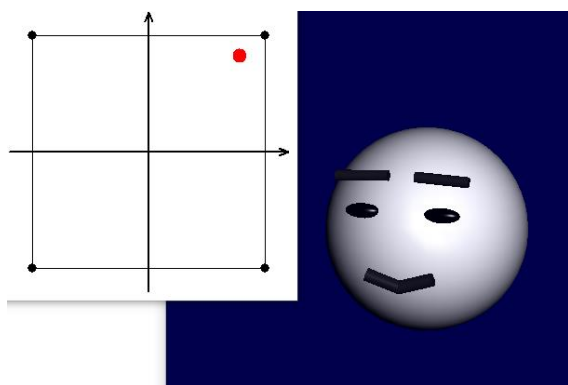


図 7 実行結果(幸福)

Fig.7 Execution result.

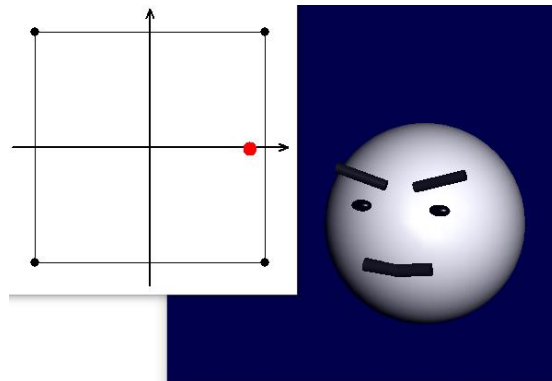


図 8 実行結果(幸福と怒りの中間)

Fig.8 Execution result.

は原点をポイントしており、図 7 は幸福（喜）の付近をポイントした結果である。図 8 は幸福と怒りの中間地点をポイントしており、2つの感情が入り混ざった表情を表している。

4. まとめ

本稿では顔情報の多次元パラメータを人間が直観的に認識できる次元で操作可能なしくみを作成し、リアルタイムに CG で顔表情を創出する UI および表情合成システムの開発を試みた。本システムは動作のしくみがシンプルであり、処理の負荷が小さいので、同時に多数の顔表情を処理するようなシステム、例えば教育支援システムや遠隔教育システムなどや、通信帯域の乏しい情報通信環境におけるコミュニケーションシステムなど、幅広い分野への応用が期待できる。

参考文献

- [1] 宮崎 光二, 中山 弘隆: 似顔絵における表情作成について, 計測自動制御学会論文集, 第 38 巻, 第 5 号, (2000).
- [2] 楠見 孝, 米田 英嗣, 小島 隆次: アバターの感情表出機能によるマルチユーザ仮想空間コミュニケーション・システムの改良, 日本教育工学会論文誌, Vol.31, No.4, (2008).
- [3] 吉田賢史, 宮崎光二, 中上香代子, 中山弘隆: e-Learning における適応的コンテンツ配信制御, (独)メディア教育開発センター, 教育メディア研究, 第 2 巻 第 1 号, No.2, pp.81-pp.91, (2005)
- [4] J.A.Russell, A circumplex model of affect, Journal of Personality and Society Psychology, vol.39, pp.1161-1178 (1980).
- [5] Ekman P., Friesen W. V. : ~ Facial Action Coding System, ~ Consulting Psychologist Press (1977)