

B2-1 自己組織化マップを用いた空間周波数に基づく表情認識に関する研究

福井大学 大学院工学研究科 知能システム工学専攻 進化ロボット研究室
松原 大典 (指導教員：前田 陽一郎、高橋 泰岳)

1. 緒言

近年、ロボットは病院や家庭などの我々人間の生活環境へと進出し身近な存在となってきた。特に、高齢化が進んでいる現在、福祉・介護ロボットの開発が重要視されている。このようなロボットには、人間との円滑なコミュニケーションを行うために、ロボットが人間の情動を読み取る能力も必要となる。

人間が相手に自分の意思を伝える際に、表情はバーバル(言語的)な情報に比べ、より大きな比重を占める。そのため、ロボットが人間とコミュニケーションを行う際、表情認識は重要な機能であり、様々な研究が数多くなされている [1]。小谷らは表情の変化量をニューラルネットワークへの入力として用いて認識を行い [2]、松久らはガボールフィルタを用いて画像の特徴を抽出し、学習モジュールと識別モジュールの2つを用いたことで微妙な表情の認識を行った [3]。

そこで本研究では、顔の全体的な立体配置情報が含まれる低空間周波数と、照明変動に影響されないガボールフィルタに注目し、認識画像から余分な情報を削減し、表情認識に必要な情報を抽出して表情マップとの一致率により認識を行う [4]。表情別に作成した表情マップと認識表情を全て含んだ表情マップを用いてそれぞれ認識実験を行ったので、これらについても報告する。

2. 顔要素のラベル付けに基づく自己組織化マップによる表情認識

表情は、主に顔の構成要素である眉、目、鼻、口などの形状や動きによって表出される。そこで筆者らは、表情表出の際に大きく関わる目と口に注目して、顔画像からの抽出を行い、人間の基本6表情と無表情の7表情の認識を行った。自己組織化マップ(SOM)を用いて、抽出した目と口の表出表情を求め、その組み合わせから表情の認識を行う。

表情を顔要素ごとに分解するのは、人物によらず、認識を行う表情への分類を行うためである。顔要素が表出している表情を組み合わせることで、認識表情として設定していない表情でも、認識表情として表すことができ、認識を行うことが可能となる。高次元データの可視化と分類ができるSOMを用いて、データベースとなる表情マップを作成し、顔要素が表出している表情の決定を行った [5]。表出表情の組み合わせによる表情は、研究室の20代の男子学生10名によるアンケートで、組み合わせた画像がどの表情を表出しているか、また複数の表情を表出しているか、を質問し、集計して決定した。

図1に作成した目と口の表情マップを、認識を行う基本6表情と無表情の7つの表情を示すユニット別に区分分けしたものを示す。図2に作成した表情マップ

に認識画像の顔要素を入力し、表出表情の組み合わせから行った、表情認識の一例を示す。認識実験より、無表情、驚き、恐れ、嫌悪、悲しみの5つの表情画像の顔要素は、目と口で同じ表情を表出していると認識され、喜び、怒りの2つの表情画像の顔要素は違う表情を表出していると認識された。表情ラベルの組み合わせによる表出表情とアンケート結果で得られた印象表情は、表出表情と印象表情の結果について、無表情、喜び、驚き、怒り、嫌悪の5つが両方とも一致した。恐れ、悲しみの表情は、顔要素が目と口、両方とも嫌悪と認識され、悲しみも同様に両方とも無表情と認識されるという認識結果となった。抽出した各表情の顔要素の表情マップ上での一致率の平均は、目は76.47%、口は79.53%となり口は目よりもやや一致率が高いことが分かった。

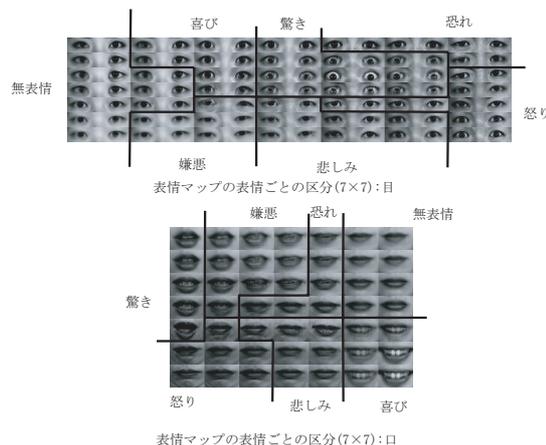


図1: SOMで作成された表情マップ

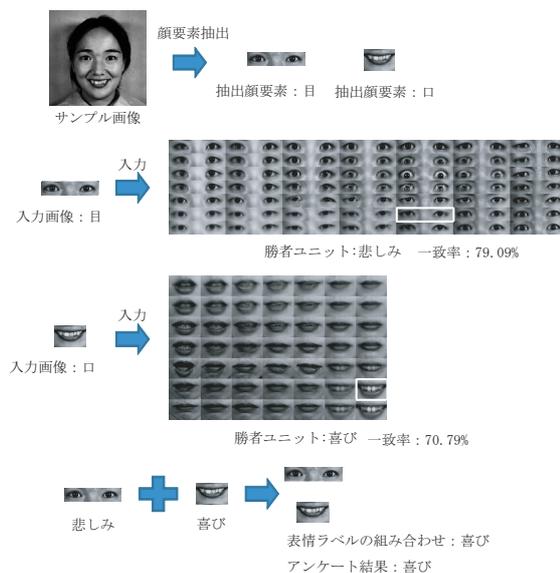


図2: 表情認識例: 喜び

3. 空間周波数とガボールフィルタを用いた自己組織化マップによる表情認識

本研究では、多様な表情認識を行うために多次元データを可視化できる SOM を用いて作成したデータベースとなる表情マップと、画像の特徴抽出に低空間周波数とガボールフィルタを用いた表情認識手法を提案する。

3.1 顔画像からの特徴抽出

表情は、顔の構成要素である顔器官の変動によって表出され、非言語情報の中でも特に重要な意味が含まれており、顔器官の形状や特徴により認識が行われる。顔器官の形状や表情表出時の変化には、個人の特徴が含まれているので、表情認識には人によらない特徴抽出が必要である。前章までの手法では、顔要素の表出表情の組み合わせによって全く違う表情として認識してしまうことがあるので、認識画像から必要がある目と口を含む顔領域全体を抽出して認識を行う。

抽出した画像から低空間周波数処理を行い、表情認識に必要な顔要素の形状情報を求める [6]。これにより背景や髪の毛などの表情の認識に関係ない余分な情報を削減し、顔要素の大まかな布置情報を求めることができる。さらに低空間周波数を抽出した画像にガボールフィルタ (以下、適宜 GF と略記) をかけて、画像の特徴抽出を行う。ガボールフィルタは 2 次元のガボール関数で、脳の第一次視覚野にある単純型細胞の空間特性を空間フィルタでモデル化でき、画像の局所的な特徴を抽出することができる [3]。また、相対的な濃淡変化を抽出するので画像の照明運動の影響を受けにくい。ガボールフィルタは x 方向と y 方向の二方向で特徴抽出を行い、認識に必要な情報を抽出することで認識を行う。図 3 に顔画像の特徴抽出の処理手順を示す。

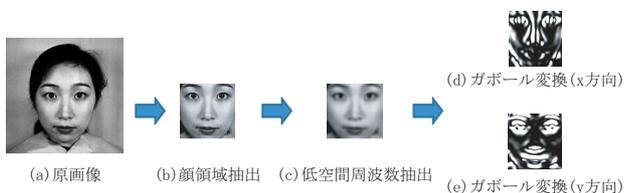


図 3: 顔画像の特徴抽出処理

3.2 表情の認識手順

表情の認識手順を図 4 に示す。本研究では、単一の表情のみで作成した各表情マップと基本 6 表情と無表情の 7 つの表情が含まれた表情マップでそれぞれ認識実験を行う。認識には表情マップの作成に用いたものとは別の画像を用いる。認識に用いる画像も表情マップの作成に用いた画像と同様の低空間周波数とガボールフィルタの処理をして、特徴抽出を行う。個別の表情画像によって作成した表情マップは、各表情マップに入力を行い、入力画像と表情マップの一致率を比較する。最も一致率が高い表情マップの表情を顔画像の表出表情として認識を行った。

ガボールフィルタの x 方向と y 方向の表情マップで表出している表情が異なった場合は、表情マップとの一致率が高い方の認識表情を表出表情とした (図 5 参照)。7 つの表情で構成された表情マップは、表情マップの各ユニットごとに表情のブロックに分け、どのブロックに認識画像が入力されたかで認識を行った。

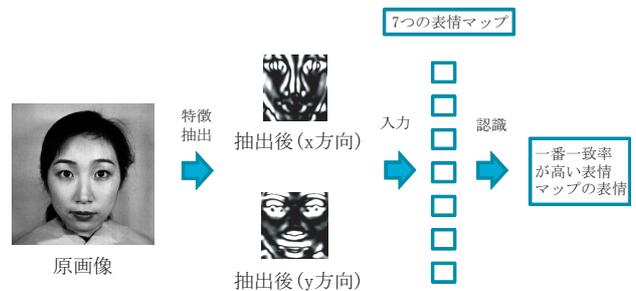


図 4: 表情認識手順



図 5: 表情認識例

3.3 単一方向および二方向ガボールフィルタによる表情認識実験

表情マップの学習回数を 5000 回とした 7×7 ユニットの 2 次元平面マップを用いて、単一方向 (x および y 方向) のガボールフィルタを用いたものと二方向のガボールフィルタを用いたものの認識率の比較を行った。顔画像からの抽出は画像内に髪の毛や首が入らないようにした。また、実験には表情マップの作成に用いた画像とは別の画像を各表情 100 枚ずつ計 700 枚を用いて行った。単一の方向の画像は x 方向と y 方向のそれぞれの認識画像を 7 つの表情マップに入力し、その中で最も一致率が高い表情マップの表情を表出表情として認識を行った。二方向の画像による認識は、単一の方向の認識と同様に各表情マップに入力して、一致率を比較して行う。認識表情が異なった場合は、表情マップとの一致率が高い方を表出表情の認識結果とする。

二方向 GF を用いた認識例を図 6 に、表情の認識率の比較を各表情別に示したものを図 7 に示す。図 6 は入力画像と x 方向、 y 方向の最も一致率が高い表情マップが示されている。表情マップの中で、入力画像と最も類似しているユニットが赤い丸で囲まれている。 x 方向と y 方向で認識表情が異なる場合、ユークリッド距離が小さい方の表情を表出表情として認識する。図 6 の場合は、認識表情が異なっているのでユークリッド距離が小さい x 方向の認識表情である喜びを表出表情とする。実験の結果、図 7 から二方向 GF の画像を用いた認識は、単一方向 GF を用いた認識よりも概ね高い認識率が得られた。単一方向の画像の認識で、驚きと悲しみの認識率が低いのは、表情表出時の特徴に方向性があるため、単一方向の GF の角度では抽出しにくく、特徴抽出をうまく行えなかったものと考えられる。

「無表情」と「喜び」の認識率は、単一方向の画像を用いた方がやや高いが、他の「驚き」「恐れ」「怒り」「嫌悪」「悲しみ」の認識率は大きく上昇し、全体の認識率を上げることができ、80% を超える認識率が得られた。 x 方向と y 方向のガボールフィルタを用いたそれぞれの認識結果より、表情には、表出時に縦と横にそ

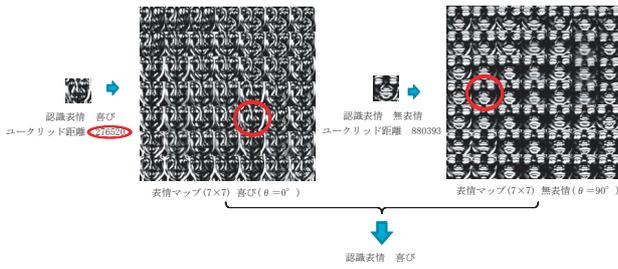


図 6: 二方向 GF 画像認識例

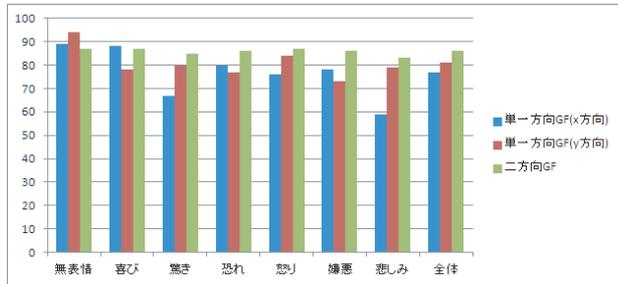


図 7: 認識率の比較 (単位%)

それぞれ大きく動いて表出するものがあり、どちらでも表出を行うものがある。「悲しみ」は縦方向に「喜び」は横方向に顔器官が動くことで表出が行われ、残りの「驚き」、「恐れ」、「怒り」、「嫌悪」は縦と横のどちらにも動いて表出が行われると考えられる。そのため、単一方向のガボールフィルタだけでなく、二方向のガボールフィルタを用いることによりどの表情の特徴も抽出できる。この結果から、二方向 GF を用いて特徴抽出を行うことで、単一方向 GF では検出しにくい表情認識も行うことができ、高い認識精度が得られることがわかった。

3.4 学習回数別の表情マップによる表情認識実験

二方向のガボールフィルタを用いた認識を、表情マップの学習回数 5000 回、10000 回、15000 回に対して行い比較した。表情マップは 7×7 の 2 次元平面マップとし、基本 6 表情と無表情の 7 つを使用したガボールフィルタ別に作成を行った。認識は特徴抽出した認識画像に方向別のガボールフィルタをかけ、同じ方向の各表情マップにそれぞれ入力し、入力した表情マップの中で最も一致率が高かった表情マップの表情を表出表情とする。 x 方向と y 方向の表情マップで、認識表情が異なった場合、表情マップとの一致率が高い方の表情を認識結果とする。一致率の比較には、最小となるユークリッド距離を用いる。最小になるということ

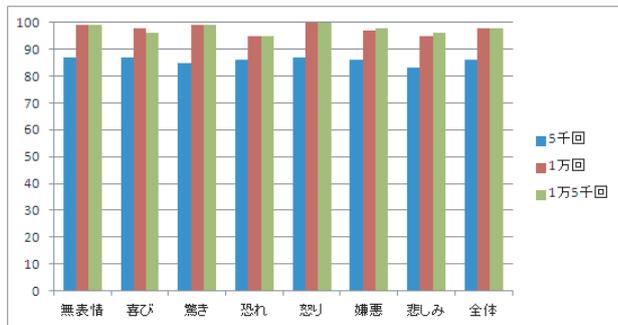


図 8: 学習回数別の認識率の比較 (%)

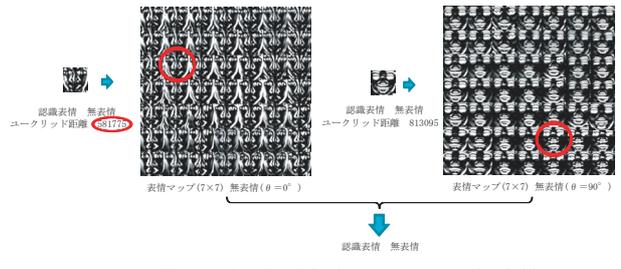


図 9: 学習回数別認識例 (5000 回): 無表情

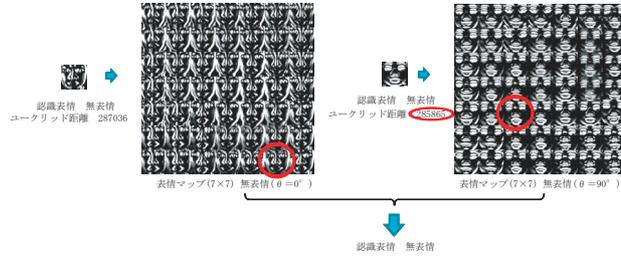


図 10: 学習回数別認識例 (10000 回): 無表情

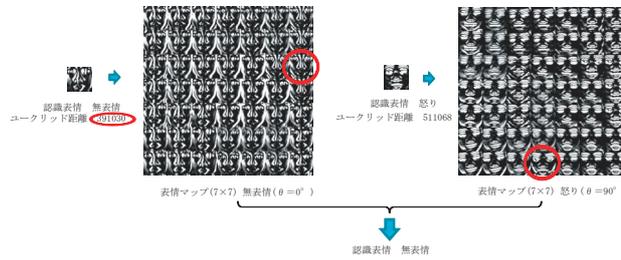


図 11: 学習回数別認識例 (15000 回): 無表情

は、表情マップと表出表情が最も類似している、ということの意味する。

学習回数ごとの認識率を比較することで、学習の収束の様子を見た。図 8 に表情マップの学習回数の違いによる認識率の比較を示す。図 8 から、学習回数が 5000 回から 10000 回に増えたところで、認識率が約 10 % 上がったことが、確認できる。15000 回の表情マップの認識率は、10000 回の学習の表情マップの認識率とほぼ同じとなったので、学習は 10000 回でほぼ収束したと考えられる。図 9 ~ 図 11 に学習回数別による表情の認識例を示す。図 9、図 10 は二方向とも無表情として認識しているため、無表情を認識表情とする。図 11 は、認識表情が無表情、怒りと異なっているため、ユークリッド距離が小さい x 方向の認識表情である無表情を認識表情とする。二方向 GF による認識を行ったことで、学習が収束していない 5000 回の表情マップでも全ての表情で認識率が 8 割を超える結果が得られた。

3.5 7つの表情を含む合成表情マップによる表情認識実験

認識を行う人間の基本 6 表情と無表情の画像がすべて含まれる表情マップを用いた実験を行った。ここでは画像から顔領域を抽出しただけのもの (実験 A)、抽出した画像に低空間周波数処理を行ったもの (実験 B)、低空間周波数処理した画像にガボールフィルタをかけたもの (実験 C) の 3 種類でそれぞれ行った。表情マップは、前節の結果より学習が収束する 10000 回で作成し、 15×15 の 2 次元平面マップとして作成する。顔画像から顔領域を抽出したもの (A) と低空間周波数処理を行ったもの (B) は、作成したマップを表情別にブ

ロックで区切り、入力された画像と最も類似したユニットとされたユニットの表情を表出表情とする。ガボールフィルタをかけたもの (C) はガボールフィルタ別に表情マップを作成し、同様の認識処理を行う。それぞれのガボールフィルタの表情マップで、認識表情が異なる場合は表情マップとの一致率が高い方の表情を表出表情の認識結果とする。

図 12 にそれぞれの手法を用いた認識率の比較を、図 13 に手法 (C) の表情認識例を示す。表情マップが表情ごとに赤い線でブロック別に区切られ、左側には入力画像が示されている。白い丸で囲まれたユニットが入力画像と最も類似した表情マップのユニットであり、そのユニットがあるブロックの表情を、画像が表出している表情とする。図 13 は、二方向とも無表情として認識されたので、表出表情は無表情と認識する。

図 12 より低空間周波数処理を行った画像にガボールフィルタをかけたもの (C) は、全体の認識率で「驚き」以外のすべての表情で他のものよりも高い認識率となり、低空間周波数とガボールフィルタを用いた手法を用いることで個人によらず、特徴抽出を行うことができたと考えられる。最も認識率が低かった「驚き」の表情は、表出時の特徴が誤認識した表情と類似していたためと考えられる。特に、最も誤認識することが多かった「喜び」とは、目や口の表出の特徴が類似しているため誤認識してしまったものと考えられる。実験の認識率を比較すると、低空間周波数処理を行ったものは、ただ画像を抽出したものよりも高く、表情表出時の形状情報を抽出することができたためと考えられる。さらに、二方向のガボールフィルタを加えたものは、他の処理よりも高い認識率となり、表情表出時の特徴を適確に抽出することができた。特に低空間周波数処理をしたことで、認識率が下がってしまった「悲しみ」の表情はガボールフィルタをかけることで、特徴抽出を行うことができた。また実験では、「喜び」などの快の印象を相手に与える表情だけでなく「恐れ」、「怒り」、「嫌悪」といった不快を与える表情の認識率も約 8 割という高い結果となり、これらの表情は他人に自分の感情を伝える重要な表情だと考えられる。

4. 結言

本研究では、表情が表出される際の特徴を低空間周波数とガボールフィルタ (GF) を用いて抽出し、データベースとなる表情マップを自己組織化マップ (SOM) を用いて特徴に応じて分類されたものを学習により作成した。表情の認識は入力画像と表情マップの一致率から行い、単一の表情だけで作成したものは、最も一致率が高かった表情マップの表情を表出表情とし、7 つ

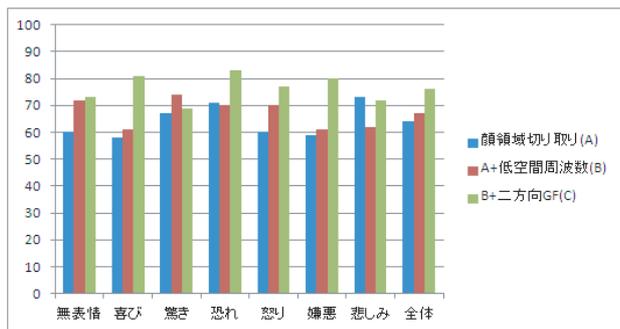


図 12: 異なる画像処理による認識率の比較 (%)

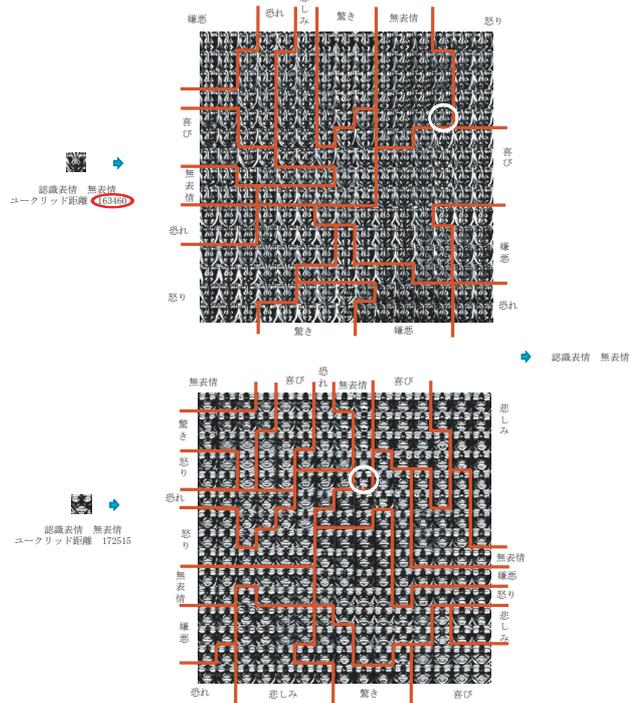


図 13: 低空間周波数 + 二方向 GF による認識例 (実験 C): 無表情

の表情を含む合成表情マップは表情別にブロックで区切り、入力された画像と最も類似したユニットとされたユニットの表情を表出表情として認識した。

今回の実験から、画像の特徴を低空間周波数とガボールフィルタを用いることで表情の特徴を抽出でき、認識を行う表情の画像が含まれる表情マップと組み合わせることで適切に表情の認識を行えることが示された。本手法を用いた実験から、ガボールフィルタによる特徴抽出を一方向だけでなく、二方向で行うことで抽出しにくい表情の特徴も抽出することが可能となり、高い認識精度になることが確認できた。

参考文献

- [1] 赤松茂, “人間とコンピュータによる顔表情認識 [II] コンピュータによる顔表情認識技術 (1): 表情による感情の認識,” 電子情報通信学会誌, Vol. J85, No.10, pp.776-771 (2002)
- [2] 小谷中洋介, 本間経康, 酒井正夫, 阿部健一, “ニューラルネットワークを用いた顔表情認識,” 東北大学保健学紀要, pp.23-32 (2004)
- [3] 松久ひとみ, 橋本学, “Gabor 特徴を用いた顔画像からの微妙な表情変化の推定,” 第 28 回ファジィ・システム・シンポジウム, pp.1143-1148 (2012)
- [4] 松原大典, 前田陽一郎, 高橋泰岳, “ヒューマンインタラクションのための空間周波数と自己組織化マップと空間周波数を用いた表情認識,” 第 28 回ファジィ・システム・シンポジウム, pp.500-503 (2012)
- [5] 石井雅樹, 佐藤和人, 間所洋和, 西田真, “自己写像特性を用いた顔表情カテゴリーの抽出と感情空間マップの生成,” 電子情報通信学会論文誌, Vol. J91-D, No.11, pp.2659-2672 (2008)
- [6] 桐田隆博, “表情の認識 -心理学の立場から-,” MEDICAL IMAGING TECHNOLOGY, Vol.12, No.6, pp.681-687 (1994)