

ニューラルネットワークによる顔の特徴点検出の評価

保黒 政大 長坂 保典
梅崎 太造 鈴木 宣夫

中部大学工学部

〒487 愛知県春日井市松本町1200

あらし

顔の識別をする際には、前処理として撮影距離や顔の位置などの正規化を行う必要がある。このときの基準点となる顔の特徴点検出の方法として、ニューラルネットワークを利用する場合について述べ、その評価をする。撮影距離や照明の明るさの違いを考慮し、ソベル変換によりエッジ強調したものを入力画像とし、特徴点で反応を示すようなニューラルネットワークを構築する。顔の特徴点として、鼻と両目について学習させて検出実験を行い、一つのニューラルネットワークのみで異なる撮影距離の顔画像の特徴点検出ができることを示し、その有効性を明らかにする。

和文キーワード 顔の識別 ニューラルネットワーク ガウス窓

The detection method for the characteristic points of human face based on artificial neural networks and its evaluation

Masahiro HOGURO Yasunori NAGASAKA
Taizo UMEZAKI Nobuo SUZUMURA

College of Engineering, Chubu University

1200 Matsumoto-cho, Kasugai-shi, Aichi 487 Japan

Abstract

It is important for recognizing human face to detect the characteristic points of human face. In this paper, the detection method and its evaluation are described which based on artificial neural networks. The edge enhanced patterns are used for the inputs of networks. The artificial neural networks can learn various size of patterns. The size of face pictures and lightning conditions are therefore not affect to detection. The efficiencies of the method to detect nose and both eyes are evaluated for the various conditions of face pictures.

英文 key words human face recognition, artificial neural networks, Gaussian window

1 はじめに

個人を同定する手法として、指紋や眼底毛細血管を利用する方法が研究され、高い成果を得ている^[1]。しかしながら、これらの方法はそれらを採用する際に特別な装置を必要とし、ヒューマンインタフェイスの点では少し問題がある。

これらに次ぐ、個人同定手段として顔画像を用いる方法が注目されている^{[2]、[3]}。顔画像を用いて個人を同定する場合、本人は特別な意識をばらわずに撮影できるため、ヒューマンインタフェイスに優れているといえる。しかし、撮影距離や顔の位置を拘束できないため、大きさや位置を正規化する必要がある^{[4]–[11]}。この際、目や鼻などの特徴点を自動的に検出し、それらを基準点とし、アフィン変換などによって正規化する。そのため、これらの特徴点は撮影条件に関係なく、正確に検出する必要がある。

これまで報告されている特徴点検出法の代表的なものとして、1) 画像を二値化し、それを濃度ヒストグラムを用いる方法^[12]、2) 特徴点のテンプレートを作成し探索する方法^{[11]、[13]}、3) カラー画像で人間の肌の色の特徴を利用する方法^{[6]、[9]}などがある。

これらの方法は、画像の濃度値をそのまま用いるため、被写体とカメラの距離や照明の明るさなどといった撮影条件が大きく影響すると考えられる。撮影距離が異なることによって特徴点の濃度分布が変化し、距離の異なる画像について一定の条件での位置検出が困難である。また、このような距離の影響についての定量的な実験報告はない。

本稿では、このような撮影条件に強く依存しない特徴点検出法としてニューラルネットワークを用いた方法を検討した。すでにニューラルネットワークを用いた位置検出法として、ナンバープレートの位置検出法が提案され高い成果を得ている^{[14]、[15]}。そこでこの方法を応用し、様々な撮影距離の顔画像に対し目や鼻の位置検出実験を行い、その評価を行った。

2 画像データベース

眼鏡を外した正面顔画像をビデオカメラで 4, 5, 6m の距離から撮影した。顔全体が画像内に入るようにのみビデオカメラの調整を行った。撮影した画像は、

縦横 512×512 画素、白黒 256 階調に AD 変換した。データの採録にあたっては特に指示を与えず、カメラの方を向いている画像を採用した。画像データ数は、男性 41 名、女性 2 名 (計 43 名) を各距離において 3 枚ずつ撮影した、総計 387 枚である。

画像データの例を図 1 に示す。



図1 画像データの例

これらの画像は、特徴点の学習と位置検出実験の評価に使用するため、あらかじめ両目と鼻の位置を視察によって調べてある。

3 学習方法

特徴点検出に使用するニューラルネットワークは三層の階層型で、学習にはバックプロパゲーション法を用いる。入力層、中間層、出力層のユニット数は各々 256, 100, 256 とする。使用したニューラルネットワークの構成を図 2 に示す。

学習は、次に示すような手順で行う。全ての出力層に対して、均一に学習を行なうため、学習パターンは学習毎に乱数を用いて切り出す位置を決定する^[16]。

- (1) 原画像から特徴点を含む 128×128 画素を乱数により切り出す。
- (2) 撮影時の照明条件の違いによる影響を考慮し、鼻等の特徴を強調するため、ソベル変換によりエッジ

ジ強調をする。ソベル変換には図3に示すような3×3のマスクを使用する。

- (3) 画像を16×16画素に圧縮し、ニューラルネットワークに入力する。教師信号は、あらかじめ視察により調べた特徴点の位置を1、その他を0とする。

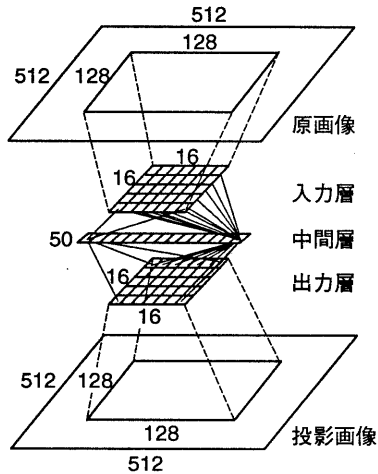


図2 ニューラルネットワークの構成

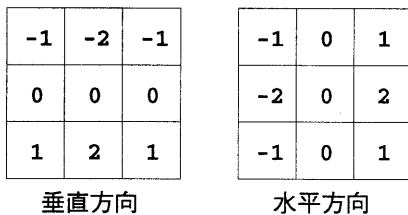


図3 ソベル変換

4 検出方法

原画像を左上から右下へ一定の幅でずらしながら切り出し、学習時と同様にソベル変換によりエッジ強調

をし16×16画素に圧縮した画像をニューラルネットワークに入力する。このようにして得られた出力値を、原画像レベルの512×512画素に投影し、加算していく。このような処理を原画像の全領域について繰り返す。最も大きな加算出力値を得たところを検出点とする(図4)。

4.1 重みなし投影加算

ニューラルネットワークへの入力画像は、切り出し画像128×128画素を16×16画素に圧縮しており、出力層の1画素は原画像では8×8画素に対応する。そこで、出力層の1画素を8×8画素に投影する。重みなし投影加算の場合、8×8画素全てに同じ重みで投影する。

4.2 ガウス窓による重みつき投影加算

全ての重みを等しくして投影すると、誤った反応を示しているところでも加算値が大きくなり、真の特徴点の値よりも大きくなる危険性が高い。そこで式(1)のようなガウス窓により中心部分に重みづけを行い、出力層の1画素を8×8画素以上の領域に投影する。このように投影することにより誤った反応をその他の正しい反応により抑えることができる。

重みなし投影加算では検出に失敗し、ガウス窓による重みつき投影加算により検出できた投影加算例を図5に示す。実線がニューラルネットワークにより検出された点、点線が実際の特徴点の位置を表している。また、ニューラルネットワークによって検出された位置での縦横1ラインの投影加算値も示した。重みなし投影加算では右目の誤った反応が強いが、重みづけをすることにより左目での反応が強くなっている。

$$g(x, y) = \exp \left\{ -\frac{\left(\frac{x}{\sigma_x}\right)^2 + \left(\frac{y}{\sigma_y}\right)^2}{2} \right\} \quad (1)$$

$$\sigma_x = \sigma_y = 1$$

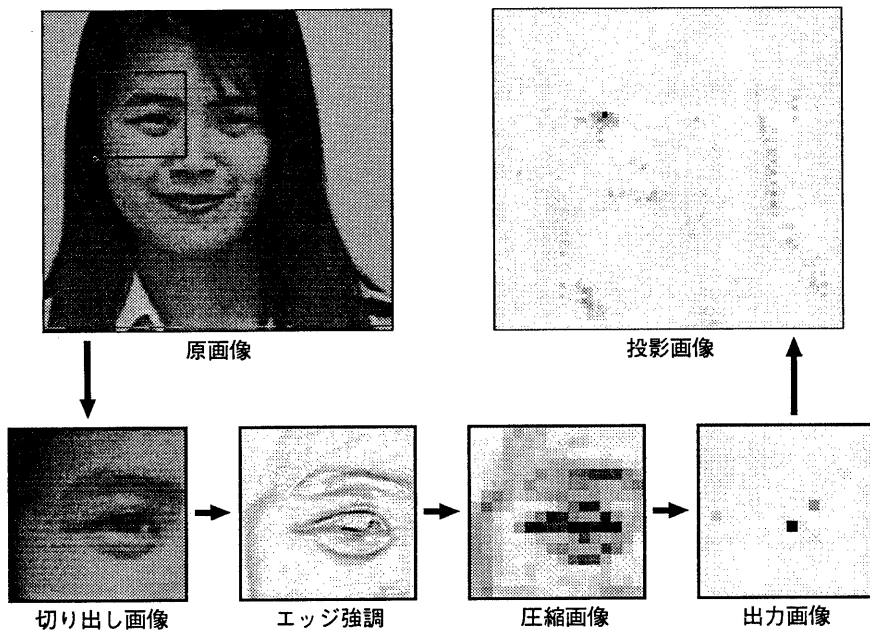


図4 検出方法

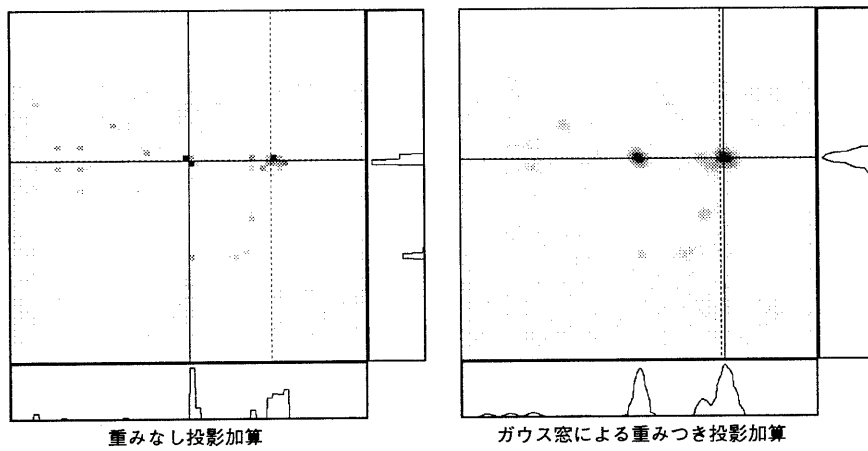


図5 投影加算の例 (左目の検出)

5 検出結果

評価実験として前述した画像データベースを使用し、鼻・右目・左目の三つの特徴点について位置検出を行った。撮影距離の異なる顔画像を混在させ、一つのニューラルネットワークのみで学習し位置検出を行った。

ニューラルネットワークの学習には、画像データベースから 22 人の各距離 1 枚ずつ (計 66 枚) を使用した。位置検出実験には、43 人の各距離で残りの 2 枚ずつ (計 258 枚) を使用した。位置検出実験に使用したデータで 22 人は学習に使用した人物と同人物であるが、学習に使用した画像データは含んでいない。

また、検出された特徴点の位置とあらかじめ視察によって求めた特徴点との距離が 8 画素以内のとき、検出成功とした。距離 8 画素というのは、出力層では 2 ユニットに相当する大きさである。今回、学習させたのは鼻の頭・目の中心の 1 画素であるため、視察によって求めた評価・学習点の精度が高くできない。このため、出力層での 2 ユニット分までは成功とした。

5.1 鼻の位置検出

鼻の頭の位置を検出する実験を行った。はじめに、重みなし投影加算をし原画像を切り出すときのずらし幅を変化させた場合の検出率を図 6 に示す。

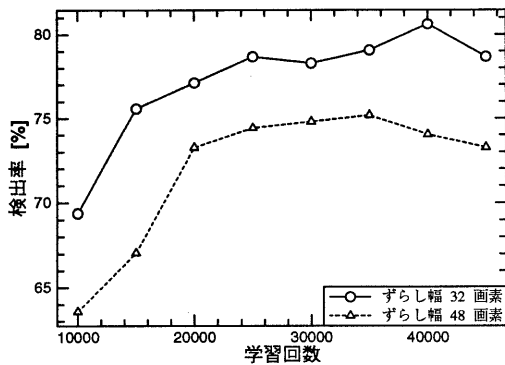


図6 鼻の位置検出率 (重みなし投影加算)

ずらし幅が小さい 32 画素のときに検出率が高く、学習回数 40000 回において 80.6% (208/258) の検出

率を得た。

ずらし幅が小さい方が出力層に特徴点の反応があらわれる入力画像が多くなるため、高い検出率が得られるといえる。しかし、ずらし幅を小さくすると検出に要する時間が増加する。そこで、今後の実験はずらし幅を 32 画素に統一した。

次に、ガウス窓を用いて投影加算した場合の位置検出率を図 7, 8 に示す。

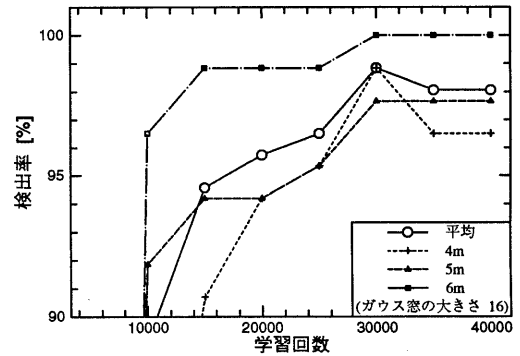


図7 鼻の位置検出率 (学習回数による変化)

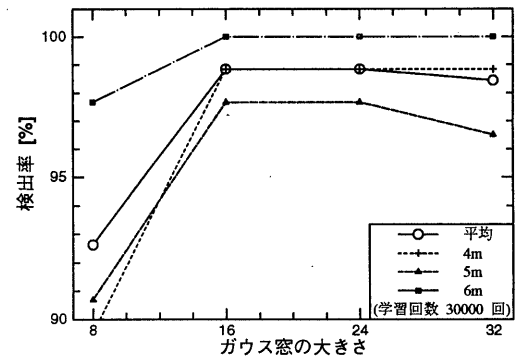


図8 鼻の位置検出率 (ガウス窓の大きさによる変化)

学習回数 30000 回、ガウス窓の大きさ 16 のときに最も高い検出率 98.8% (255/258) を得た。

検出を失敗した画像 3 枚についてみると、全て 11

画素以内に検出点を得ており、切り出すときのずらし幅を小さくすることによって検出が成功する。3枚については、ずらし幅を8画素にすることによって検出できた。また、これらの画像は未学習人物である。個人の鼻の形を考慮して学習データを選択したり、学習データ数を増やすことによって改善が期待できる。

画像の距離の違いについてみてみると4m, 5m, 6mと距離が遠ざかるにつれて検出率が高くなっている。距離が遠い画像ほど鼻の頭の部分は小さな領域となり近いと大きくなる。そのため、距離が遠い6mの画像において最も高い検出率を得たと考えられる。

5.2 右目の位置検出

右目の中心位置を検出する実験を行った。ずらし幅を32画素としガウス窓による投影加算を行った場合の位置検出率を図9, 10に示す。

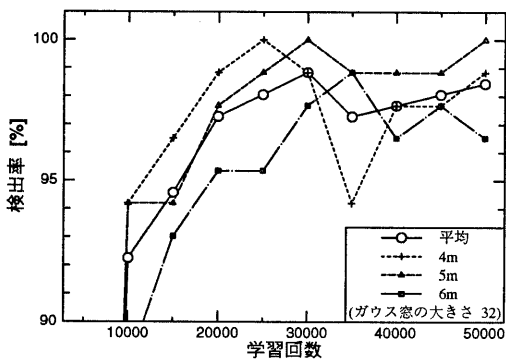


図9 右目の位置検出率(学習回数による変化)

学習回数30000回、ガウス窓の大きさ32のときに最も高い検出率98.8[%](255/258)を得た。

検出に失敗した3枚の画像は、1枚は左目に反応し、残りは顔と背景のエッジの部分に反応した。これらの画像の人物は未学習人物であり、学習データを考慮することにより検出率の向上が期待できる。また、左目に反応した画像については、鼻の位置を基準に存在領域の限定ができるため、簡単なルールの追加で検出が可能になると考えられる。

撮影距離の違いについてみると、どの距離でも良く

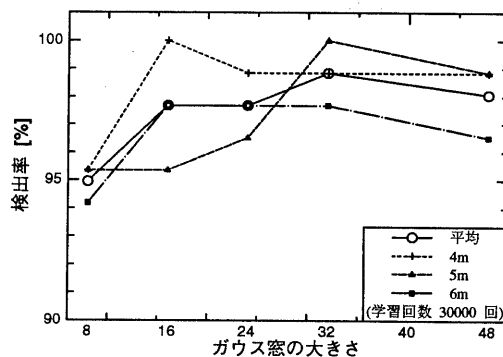


図10 右目の位置検出率(ガウス窓の大きさによる変化)

似た検出率を示している。鼻に比べ目の場合は変動が少なく、距離による影響が小さいと考えられる。

5.3 左目の位置検出

左目の中心位置を検出する実験を行った。ずらし幅を32画素としガウス窓による投影加算を行った場合の位置検出率を図11, 12に示す。

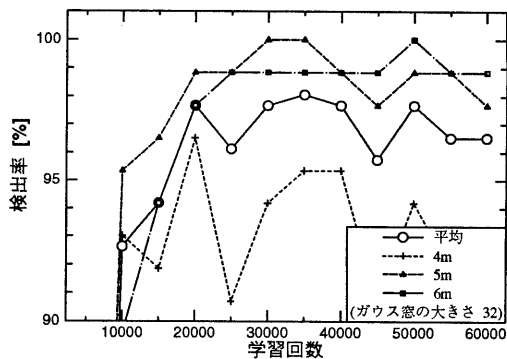


図11 左目の位置検出率(学習回数による変化)

学習回数35000回、ガウス窓の大きさ32のときに最も高い検出率98.0[%](253/258)を得た。

検出に失敗した画像5枚についてみてみると2枚が右目に反応しており3枚がそれ以外の部分に反応

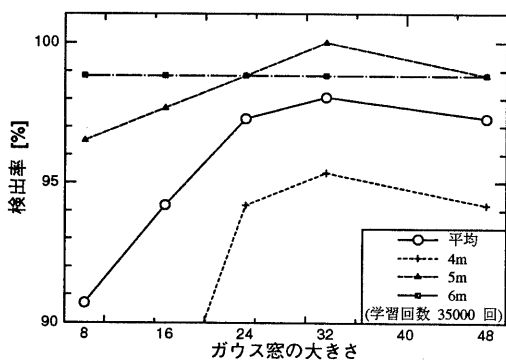


図12 左目の位置検出率 (ガウス窓の大きさによる変化)

している。この3枚は未学習人物であり、学習パターンを考慮することによって検出できると考えられる。また、他の2枚については右目同様、簡単なルールを追加することにより検出できるようになると考えられる。

画像の距離の影響は4mの画像に対して極端に検出率が悪いが、これらも右目に反応したり未学習人物の影響が特に強く出たと考えられ、これらの影響を考慮することにより検出率の向上が期待できる。

6 むすび

顔の識別をする際に必要な顔の特徴点検出について、ニューラルネットワークを用いる方法について評価実験を行った。ニューラルネットワークの出力値を、適度に調整された大きさのガウス窓を使って投影することによって、鼻と右目で平均98.8[%] (255/258)、左目で平均98.0[%] (253/258)の位置検出率を得た。平均してみると、全ての距離の画像に対して高い検出率を得たといえる。

一つのニューラルネットワークにことなる距離の顔画像を学習させ、位置検出を行った。ニューラルネットワークを用いると、距離の異なる画像に対して適したテンプレートが自動的に学習・選択されたと考えられる。各距離別に検出率を見てみると、全ての距離において高い検出率を得ており距離の違いに柔軟に対応し

た位置検出ができたと言える。また、ニューラルネットワークに入力する画像はソベル変換によりエッジ強調した画像であり、照明条件の影響を受けにくいシステムであると考えられる。

検出を誤った画像について考察すると、右目を左目、あるいは左目を右目と間違えていることがある。今回の実験では鼻・右目・左目の検出を独立に行っているが、一般的な相互の位置関係を考慮することによってさらに検出率の向上が期待できる。また、全く違うところに検出点を示した場合は、それらの画像は学習に使用しなかった人物ばかりであり、鼻等の形の個性が影響していると考えられる。これらの画像についても学習する人物を検出し、増加することにより検出率の向上が期待できる。

参考文献

- [1] B. Miller: "Vital signs of identity", IEEE Spectrum, February (1994).
- [2] 南敏: "顔の識別技術", 計測と技術, **25**, 8, pp. 707-713 (1986).
- [3] 南敏: "顔画像による個人識別の技術", システム制御学会誌, **35**, 7, pp. 415-422 (1991).
- [4] 赤松茂, 佐々木努, 末永康仁: "正面顔画像による個人認識の基本検討", テレビジョン学会技術報告.
- [5] 赤松茂, 佐々木努, 深町映夫, 末永康仁: "KL展開によるパタン記述法の顔画像識別への応用", 電子情報通信学会研究報告, PRU90-152, pp. 55-62 (1990).
- [6] 赤松茂, 佐々木努, 深町映夫, 末永康仁: "濃淡画像マッチングによるロバスタな正面顔の識別法—フーリエスペクトルのKL展開の応用—", 電子情報通信学会論文誌, **J76—D—II**, 7, pp. 1363-1373 (1993).
- [7] 小杉信: "モザイクとニューラルネットワークを用いた顔画像の認識", 電子情報通信学会論文誌, **J76—D—II**, 6, pp. 1132-1139 (1993).

- [8] 佐々木努, 赤松茂, 深町映夫, 末永康仁: “正面顔画像の自動識別法の検討”, 電子情報通信学会報告, IE91-50, pp. 1-8 (1991).
- [9] 佐々木努, 赤松茂, 深町映夫, 末永康仁: “顔画像認識のための色情報を用いた顔の位置合わせ法”, 電子情報通信学会研究報告, IE91-2, pp. 9-15 (1991).
- [10] 島田聡, 小池秀樹, 伴野明, 石井健一郎: “顔の向きによらない人物識別法の評価”, 電子情報通信学会研究報告, PRU93-79, pp. 23-30 (1993).
- [11] D. J. Beymer: “Face Recognition Under Varying Pose”, A.I Memo, No.1461, Artificial Intelligence Laboratory, Massachusetts Institute of Technology (1993).
- [12] 黒田英夫, 光田和臣, 池本雄一郎, 藤村誠, 池原雅章: “顔画像からの口部領域の自動抽出法”, 電子情報通信学会研究報告, IE91-3, pp. 17-24 (1991).
- [13] 長谷川浩史, 那須靖弘, 清水英二: “多重解像度画像処理 (全域的, 局所的) を用いた顔画像の特徴点抽出の一方法”, 電子情報通信学会研究報告, PRU89-26, pp. 57-62 (1989).
- [14] 藤吉弘亘, 梅崎太造: “ナンバープレートの位置検出法の評価”, 電子情報通信学会研究報告, PRU93-40, pp. 45-52 (1993).
- [15] 藤吉弘亘, 梅崎太造: “ニューラルネットワークを用いたナンバープレートの位置検出—学習パターンの提示方法に関する検討—”, 電気関係学会東海支部連合大会, 741, p. 371 (1994).