

ニューラルネットワークを用いた鼻位置の検出法とその評価

菅井 満*, 藤吉 弘亘*, 梅崎 太造**, 吉田 博明***

Detection Method of the Nose Position Using the Artificial Neural Networks and its Evaluation

Mitsuru SUGAI*, Hironobu FUJIYOSHI*, Taizo UMEZAKI** and Hiroshi YOSHIDA***

(Received May 31, 1997)

The purpose of this study is to measure the degree of a person's weariness using correlation between the person's weariness and the temperature variation of his/her nose surface. As a first step a method of tracking the nose position in input images correctly in real time was developed. It is possible to detect the memory quantity with a small capacity, high speed and high precision because it uses artificial neural networks (N.N) for the position detection. In the evaluation experiment, the position detection rate of 100[%] was achieved with 266 front face images. Also, it show that it is possible to detect a nose position of a turned face by learning from some turned facial images.

Keywords : position detection, artificial neural networks, weariness measurement

1. まえがき

原子力発電所等の重要施設において、人的操作ミスによる事故は絶対に防止しなければならないが、疲労が蓄積して集中力が低下するとミスが多発する要因となる。このため、自動的に実時間で疲労度を計測する手法が関係業界より求められており、その一つとして、人の鼻の表面温度変化と疲労度の相関を利用する手法がある¹⁾²⁾。オペレータの顔を常時監視し、鼻の温度を測定することで疲労度をチェックできるが、温度測定点である鼻のトラッキングが不可欠であり、顔の上下・左右の移動や回転に追従できることが要求される。そのため、鼻の位置を顔の向きに依存せずに取得する方法が必要となる。

顔の特徴部位には、目・鼻・口等の器官が挙げられる

が、個人により大きさ・形状は多少異なる。また、表情変化による変形も生じるため、これらを吸収し、不特定の顔に対して正確に特徴部位の検出を行わなければならない。

これまでの代表的な手法として、テンプレートマッチング法が挙げられるが、不特定の顔画像に対応するためには、テンプレートを数多く用意しなければならない、記憶容量が増大する³⁾⁴⁾。宋らの手法は、外形変化の微小な特徴をテンプレートとして持つことで、個人性を排除したものであるが、二値化画像への適用であるため、閾値設定による影響が大きい⁵⁾。

また、黒田らによる二値化画像のヒストグラム法は、画像の濃淡情報を用いるため比較的良好な結果が得られるが、撮影距離や顔の向き等の影響を受け易いという問題がある⁶⁾。

本研究では、人間の疲労度を測定するために、特徴部位として鼻に注目し、ニューラルネットワーク(以後N.

*電気工学専攻大学院生(Graduate Student, Electrical Engineering)

**電子工学科(Department of Electronic Engineering)

***三洋電機メカトロニクス研究所(Mechatronics Research Center, SANYO Electronics Co., Ltd.)

Nと表す)を使用することで, 顔の向きや撮影条件等に依存しにくい特徴点検出法を提案する。また, 評価実験により本手法の有効性を証明する。

2. 鼻の位置検出

2.1 ネットワークモデル

鼻の位置検出には, 3階層型のN.Nを使用する⁷⁾。学習アルゴリズムとしては, 誤差逆伝搬学習則 (バックプロパゲーション)を用いる⁸⁾。ネットワーク中の任意の神経ユニットにおける出力を y , 入力 x とすると, その動作は次式にて表される。ここで, β はニューロンのオフセットである。神経ユニットは, 神経細胞の最小単位で, 複数結合することによりN.Nが構成される。

$$y = f(x + \beta) \quad (1)$$

ユニットの閾値関数 f としては, シグモイド関数を用いる。ここで, u は関数への入力値を表す。 U_0 はシグモイド関数の傾きで, 神経の応答特性を調整する役目を持つ。

$$f(u) = \frac{1}{1 + \exp(-\frac{u}{U_0})} \quad (2)$$

評価実験に使用するネットワークモデルを図1に示す。

2.2 N.Nの学習方法

以下に, 学習を行うまでの処理手順を示す。

(1) 画像切り出し

原画像より128×128画素のサイズで鼻が含まれる領域を切り出す。鼻が中心位置にきた場合, 個人差はあるが両目・口・唇の一部が切り出し領域に含まれるように考慮してある。なお, 切り出し位置は乱数により決定する。

(2) エッジ強調

N.Nに入力する前処理として, 8近傍のsobelフィルタを使用し, エッジ強調を行う⁹⁾。顔と背景間の輝度差が大きいため, 主に顔の輪郭が強調され易い。

(3) 平滑化

エッジ強調された目, 鼻, 口唇及び顔輪郭等の部分には個人性情報が含まれているため, そのままN.Nに入力すると過剰学習が起り, 特定の顔にのみ反応する神経が構築される。故に, 個人性情報を低減し, “顔らしさ”を表現するために, 平滑化処理を行う必要がある。

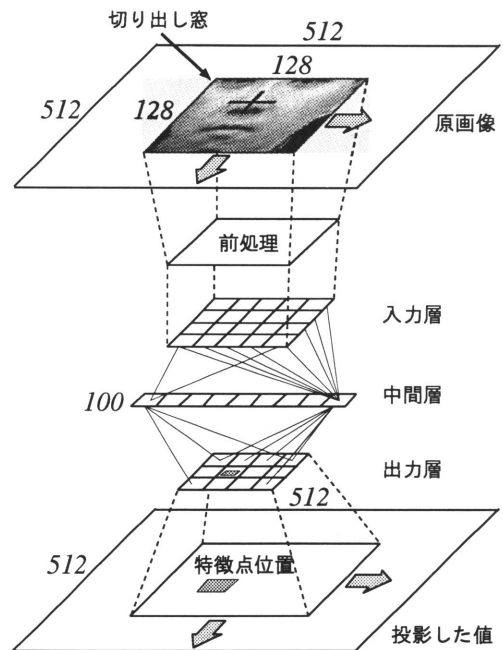


図1 ネットワークの構造

(4) 教師信号

入力パターンと教師信号の例を図2に示す。入力画像中の鼻位置に反応するN.Nを設計するため, 教師パターンは, 視察により定めた鼻位置のみを1, その他を0とする(図2(d)参照)。なお, これらの図は128×128画素を16×16画素に圧縮した例である。

(5) N.Nへの入力

N.Nの入力層に圧縮パターン, 出力層に教師パターンを提示し, 誤差逆伝搬学習則にて学習を行う。ここで, 最大学習回数は50000回に設定してある。図3に学習誤差曲線の例を示す。ユニットの閾値関数であるシグモイド関数の傾きを変えることで, 誤差収束の速度や収束性が大きく変化する。

2.3 位置検出

N.Nを用いて, 鼻の位置検出を行う手順を以下に示す。

(1) 2.2節と同じ手順で, 原画像より切り出した画像の平滑化を行い, N.Nに入力して出力値を計算する。

(2) N.Nの出力サイズは, 入力パターンと同サイズであるため, 原画像レベルの128×128サイズに拡大して

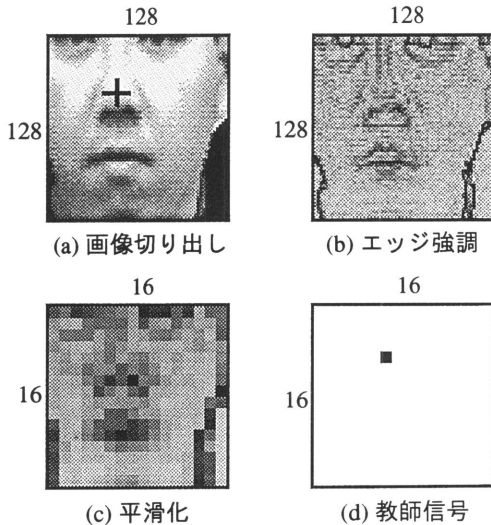


図2 入力パターンと教師信号の例

投影加算する。

- (3) 図4のように、探索窓を x, y 両方向に一定の画素(ここでは16)ずつシフトしながら探索領域全体を走査する。その間(1)~(2)の処理を繰り返す。
- (4) 投影加算した中で、最も出力値が大きい位置を鼻位置とする。

視察により求めた鼻位置とN.Nが出力した位置のユークリッド距離により検出判定を行う¹⁰⁾。視察位置を (x_0, y_0) 、N.Nの出力位置を (x', y') とすると、2点間のユークリッド距離は次式で表される。

$$d = \sqrt{(x_0 - x')^2 + (y_0 - y')^2} \quad (3)$$

入力パターンの圧縮倍率が $1/8$ である場合、 $d \leq 8[\text{pixel}]$ のとき検出成功とする。これは、N.Nの出力層の1画素が、入力画像の8画素に相当するためである。

3 顔画像データベース

使用する顔画像は、赤外線カメラとビデオカメラで同時に撮影し、ビデオ画像に赤外線画像をマスキングしたものである。そのため図5のように、顔皮膚以外の領域である背景、髪等の部分では低温のため、暗階調で表現される。

画像の大きさは $512 \times 434[\text{pixel}]$ で、白黒256階調である。ファイルサイズは222[kbytes]である。学習、評価実

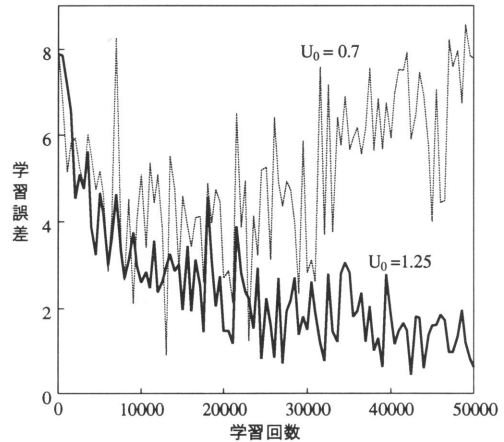


図3 学習誤差曲線

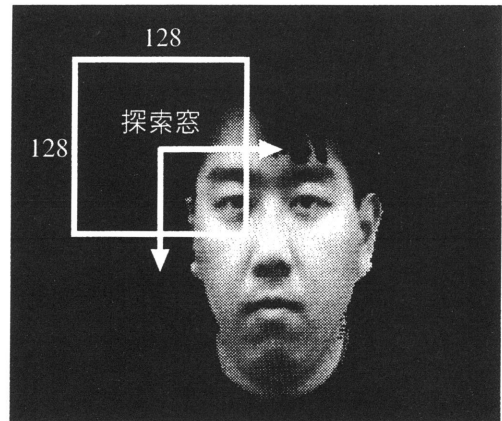


図4 鼻位置の探索方法

験に使用する画像には、撮影距離による照明の輝度差を補正するため、濃度値の変換を行う。

データベースには、正面の他に上下方向、水平左右方向(3次元的回転)に回転させた顔を撮影する。回転角度は $0^\circ, 22.5^\circ, 45^\circ$ である。一名分の顔画像の内訳を表1に示す。

実験で使用する顔画像は、回転角度により分類する。 0° グループは回転していないため、正面顔のみが含まれることになる。画像枚数は 0° の正面顔が266枚で、 22.5° 及び 45° 回転した顔画像が共に152枚ずつある。図6に回転した顔画像を示す。撮影人数は男性26名、女性12名の計38名であり、全画像枚数は570枚となる。

4 正面顔画像による実験

4.1 圧縮倍率の検討

2.2節に示したように、原画像より切り出したパターンに平滑化処理を施す。ここで、平滑化の度合いが過度に大



図5 入力顔画像の例

表1 1名分の顔画像データ

回転方向	枚数
正面	3
上	3
下	3
左 (水平)	3
右 (水平)	3

きいと個人情報だけでなく、“顔らしさ”の情報まで除去することになり、鼻の検出能力に影響を及ぼす。そこで、最適な平滑化を正面顔画像のみを用いて実験的に求める。学習には、無作為に抽出した男性5名と女性3名分の16枚（1名につき2枚）の画像を用いる。評価実験には38名分の266枚を使用する。平滑化の度合い、画像の圧縮倍率と密接な関係があることを利用する。切り出し時128×128 [pixel]の大きさを32×32～8×8サイズまで変化させて学習し、検出率の推移を調べる。なお、N.Nの入出力層は圧縮後の画像と同サイズとする。

実験結果を図7に示す。圧縮率を上げることで、検出率もある程度まで上昇傾向を見せる。原画像を1/8に圧縮した16×16サイズのとき、94.0(250/266) [%]の最も高い検出率が得られた。圧縮率を上げることで顔画像にぼかしが入り、顔の個性が失われて行く。本実験より圧縮率1/8が最適であることが判明した。

今回は、圧縮法として単一の手法のみを使用したため、この圧縮方法が最適かどうかは不明である。今後、複数の圧縮法を評価検討することが必要である。

4.2 学習領域の検討

鼻位置の検出時、実際の位置は未知であるため、入力

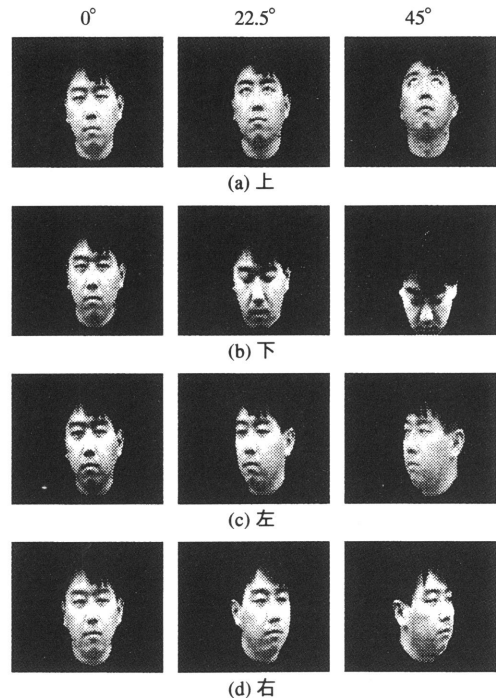


図6 回転角度による顔の変化

原画像を座標左上から右下まで全探索する必要がある。また、学習時には、鼻の位置が画像全体に分布するように考慮しなければならない。そこで、切り出し窓内で鼻の位置を乱数を用いて変動するようにする。切り出し窓の中心位置を(0, 0)と定義するとき、30～60画素のユークリッド距離範囲で、図8のように学習位置を変動させる。学習領域に対するN.Nの検出能力を図9に示す。ここで、圧縮倍率は1/8、検出判定距離は8画素とする。学習及び評価用画像は4.1節と同じである。

半径方向54画素の範囲で位置を変動させた場合、98.9(263/266) [%]の検出率が得られた。このとき、50 [%]以上の顔面積が切り出し領域に含まれることになる。

なお、本実験で用いた切り出し窓のサイズは128×128であるが、今後さらに最適なサイズについても検討する必要がある。

4.3 顔の左右非対称性

人の顔は、厳密には左右非対称であるため、実写した顔画像をミラーリングして得られた画像も実験用画像として利用でき、見かけ上のデータ数を増やすことが可能である。また、左右への顔の振れに対して一枚の画像で対応できるというメリットもある。

図10に実画像とミラーリング画像の表示例を示す。表2は、学習用顔画像にミラーリング画像も加味した場合の

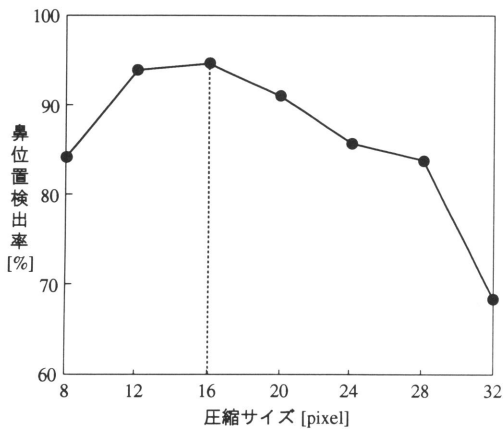


図7 圧縮サイズに対する検出率の変化

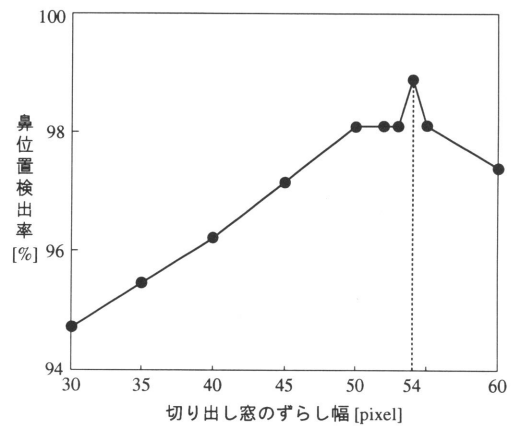


図9 学習領域に対する検出率の変化

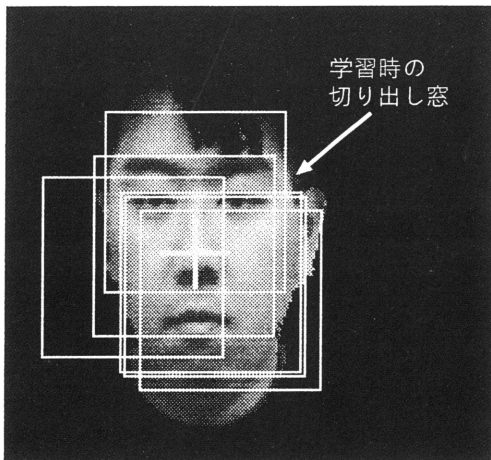
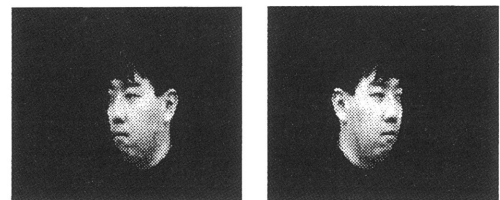


図8 切り出し窓の位置変動例



(a) 実画像 (b) ミラーリング画像

図10 ミラーリング画像例

表2 ミラーリング画像を加味した結果

学習データ	検出率 [%]	平均距離 [pixel]
実画像, ミラー 16	99.2 (264/266)	3.7
実画像, ミラー 8	96.2 (256/266)	14.4
実画像 16	98.9 (263/266)	5.7

実験結果である。16名分の正面顔のみと8名分のミラーリング画像を含めた16枚で学習させた場合では、前者の方が検出率が高い。これは、顔の変動性をN.Nが十分に吸収できないためと考えられる。一方、32枚で学習した場合、検出率そのものにほとんど差はないが、鼻位置の検出精度が約2画素分向上しており、ミラーリング画像の有効性が示された。

4.4 画像階調の変換

今回用いた画像は、通常の室内照度で撮影されたもので、バックライトは使用していない。これは実際のオペレータ監視状況を考慮したためであるが、そのため顔領域の明度が一定でなく、画像により平均階調が最高で約39 [%]変化している画像も含まれている。階調を任意に変換した画像を学習に用いることで、撮影時の照度変化を吸収できると考えられる。原画像より切り出したパター

ンに階調変換を行う。変換には次の4手法を用いる。

- 変換せず
- パターン中の階調分布を8ビットに正規化
- パターン中の最低階調値が0になるよう分布をシフト
- パターン中の最高階調値が255になるよう分布をシフト

階調値変換した画像例を図11に示す。

これらの画像を加えて学習させたが、階調値変換をしないときと比較して、検出率、精度共に大幅な向上は見られない。評価用画像には、顔領域の明度が極端に明るいもの、または暗いものは含まれておらず、学習効果が十分に発揮されていないと考えられる。



図11 階調変換例

表3 階調変換画像を付加した結果

学習データ	検出率 [%]	平均距離 [pixel]
階調変換有り	99.2 (264/266)	3.8
無し	99.2 (264/266)	3.7

4.5 抑制学習

入力画像中の目的部位以外に応答しないN.Nを設計する場合、目的部位に類似したパターンや小領域しか目的部位付近の特徴が少ししか含まれていないパターンに対しては、抑制学習が有効である¹⁰⁾。顔の領域が切り出し時にどの程度含まれるかを示す“切り出し率”を定義し、この割合以下で切り出された画像は抑制パターンとして用いることを考える。切り出し率の算出は、次のように行う。 W, H をそれぞれ切り出し領域の横幅、縦幅とすると、その面積 S_w は次式で表される。

$$S_w = WH \quad (4)$$

顔領域を長方形で近似した場合、左上端点と右下端点の位置をそれぞれ (x_{ul}, y_{ul}) , (x_{lr}, y_{lr}) とすると、顔の面積 S_f は次式で求まる。

$$S_f = |x_{lr} - x_{ul} + 1| |y_{lr} - y_{ul} + 1| \quad (5)$$

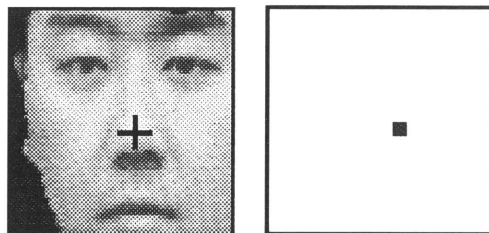
故に、切り出し時に含まれる顔面積 S_s は、次式で示される。

$$S_s = S_w \cap S_f \quad (6)$$

S_s の S_f に対する割合、すなわち切り出し率 R_{face} は次

入力パターン

教師パターン



(a) 興奮パターン

入力パターン

教師パターン

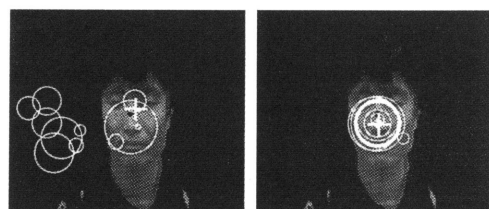


(b) 抑制パターン

図12 学習パターン例

表4 抑制学習を加味した結果

切り出し率 [%]	検出率 [%]	平均距離 [pixel]
20	99.2 (264/266)	3.6
30	99.6 (265/266)	3.7
40	100.0 (266/266)	3.6



(a) 抑制無し

(b) 抑制有り

図13 抑制学習による効果

式となる。

$$R_{face} = \frac{S_s}{S_f} \times 100 \quad (7)$$

図12に学習パターン例を示す。

切り出し率を20, 30, 40[%]と変化させて、抑制学習を付加したときの検出実験を行う。興奮学習は4.1節と同様である。

実験結果を表4に示す。切り出し率の増加と共に検出率が向上している。40[%]の切り出し率で、100.0[%]の検

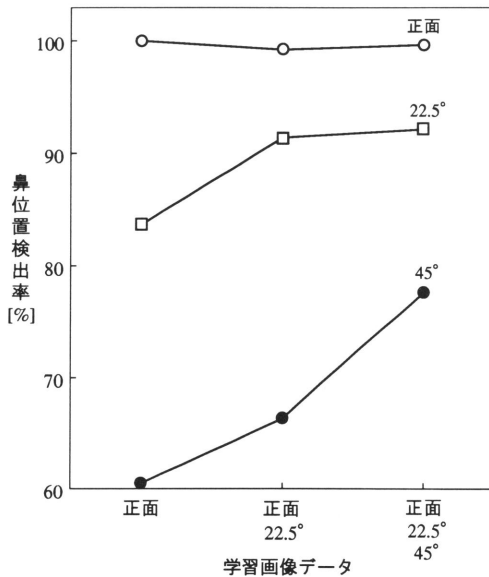


図14 正面・回転顔を学習した N.N による検出結果

出率が得られた。前の実験では、鼻筋や目頭付近への誤りが多く見られたが、抑制学習を加えることにより、検出可能となった。これにより、正面顔に関しては高精度に位置検出できることが判明した。検出例を図13に示す。白線の円とクロスマークは、それぞれN.Nの反応強度及び検出位置を示している。N.Nの反応が鼻に集中しており、背景への誤反応が無いことがわかる。

5 回転顔画像による実験

5.1 正面顔のみ学習する場合

これまでの実験では、学習と評価実験に正面顔のみを用いてきた。ここでは、正面顔のみで学習したN.Nを用いて、回転した顔画像に対してどの程度対応できるか調査する。評価用データとして、22.5°及び45°回転した顔画像を用いる。なお、0°回転を正面画像とする。

実験の結果、22.5°回転顔で83.6(127/152) [%], 45°で60.5(92/152) [%]の検出率が得られた(図14参照)。回転した顔に対してある程度対応できるが、高精度な検出は望めないことが明らかである。

5.2 回転顔を加えて学習する場合

顔の回転に対処するため、学習データとして、①正面+22.5°, ②正面+22.5°+45°の2種類のデータベースを作り、N.Nの学習に用いる。左右に回転した顔画像は、それぞれ一人につき2枚、計8人分(16枚)を用いる。選択する人物は先の実験と同一人物である。評価実験方法は5.1節と同様に行う。

上記の条件で学習したN.Nの評価実験結果を図14に

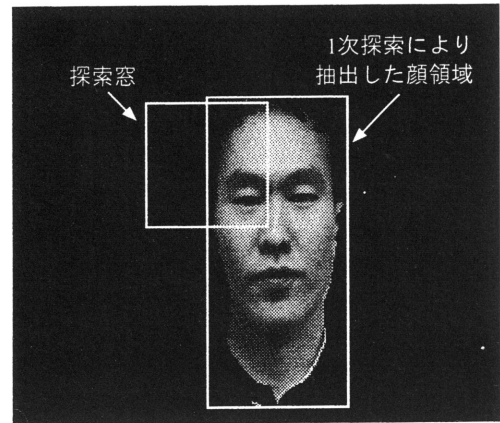


図15 1次探索による顔領域の抽出例

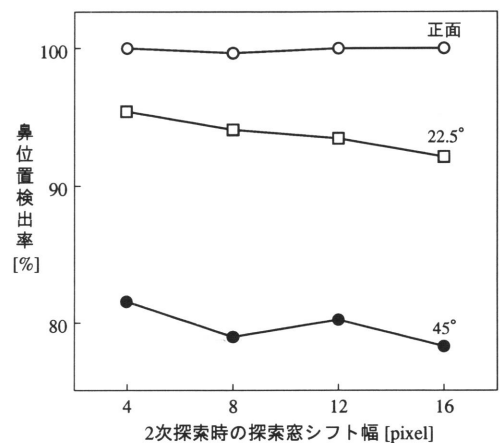


図16 1次探索の探索窓シフト幅 [pixel]

示す。正面顔99.2(264/266) [%], 22.5°回転顔92.1(140/152) [%], 45°回転顔77.6(118/152) [%]と最も高い検出率が得られており、2つの回転顔画像を学習時に付加したことが有効に作用している。位置検出に失敗した画像のうち、大多数が下・左・右を向いている画像である。上を向いている画像の場合、目、鼻、口等の隠れや変形が少なく、回転角が増加しても正面顔に近い形態が保持されているため、急激に検出率が低下することは少ない。

6 探索領域の限定

6.1 顔領域の抽出

人間の鼻位置を高速に検出するには、探索領域を顔のみに限定し、背景部分への無駄な探索を減らすことが必要である。本実験で使用している画像は図15に示すように、顔領域のみが明階調で表現されている。そのため、画像全体の階調分布を調査することで、顔領域を抽出す

ることが可能である。本稿では、画像中より顔領域を抽出する過程を1次探索とし、抽出した結果を用いて鼻位置を検出するまでを2次探索とする。

顔領域の抽出手順は、入力画像の x , y 方向それぞれについて濃度ヒストグラムを計算し、閾値により切り出した長方形領域を顔部分とする。

2次探索領域は、1次探索より求まる領域を x , y 方向にそれぞれ64画素、32画素だけ拡大したものである。一般に人の顔は縦長であり、横方向のスキャンニング回数を増やすために、 x 方向のマージンを多くしている。

6.2 1次探索の効果

1次探索の有効性を調査するために、評価実験を行う。実験に用いるN.Nは5.2節で作成したもので、正面・回転顔(22.5°, 45°)画像を学習させている。2次探索での探索窓シフト幅を4~16[pixel]の範囲で変化させ、検出率の推移を調べる。

探索窓シフト幅を4[pixel]とした場合、正面顔で100.0[%], 22.5°回転顔で95.4[%], 45°回転で81.6[%]の検出率が得られた。シフト幅が少量である程、顔領域を探索する回数が増加し、検出精度は向上するが、逆に検出に要する時間は増加する。画像1枚当りの検出時間は、Alpha21164/500[MHz]CPUの端末を使用した場合、シフト幅16[pixel]が3.2[s]であるのに対し、4[pixel]では約10倍の30[s]となる。検出率と検出に要する時間とはトレードオフの関係にあるため、これらの妥協点を見つけることが課題である。

7. むすび

本稿では、N.Nを用いて顔の向きに影響されにくい鼻の位置検出法を提案し、評価実験を通してその有効性を示した。

学習時、顔画像の切り出しに、乱数を使用することで基本的に同じパターンが再び入力されないように工夫しており、N.Nの汎化能力を高めている。また、画像データを裏返しにしたミラーリング画像を使用することにより、見かけ上のデータ数を増やすとともに左右方向への顔の回転を同時に学習できることを示した。さらに、学習データに含まれる顔面積を考慮して抑制学習を行うと、位置検出の精度が向上した。

これらのパターン提示法で学習した結果、正面顔の画像266枚に対しては100[%]の検出率を得ることができたが、回転した顔に対しては急激に検出率が低下することが判明した。

そこで、学習時に回転顔画像を加えることで、22.5°回転の画像に対して92.1[%], 45°回転に対して77.6[%]の検出率を得ることができた。

また、画像中の顔領域を抽出し、探索範囲を限定する1次探索を加味することで、22.5°, 45°回転顔において、それぞれ95.4[%], 81.6[%]と検出率の向上を図ることができた。

本手法を用いることで、従来困難であった正面以外の顔に対しても、ある程度対処できることを示した。今後さらに高精度な検出方法について検討して行きたい。

謝 辞 本研究は通産省の産業科学技術開発プロジェクト「人間感覚計測応用技術の研究開発」の委託研究の一環として実施したものであり、関係各位に感謝致します。また日頃より、本研究に関して有益な助言を頂いている三洋電機(株)メカトロニクス研究所ヒューマンサイエンス研究室の皆様にも深く感謝致します。

参 考 文 献

- 1) 菊本 誠, 吉田倫幸, 源野広和, 松本和夫, “連続トラッキング作業に伴う鼻部皮膚温の変動”, 日本人間工学会誌, 第31巻, 特別号, pp.392-393(1995).
- 2) 石川恵子, 菊本 誠, 源野広和, 鈴木龍司, 大隅正人, “顔面皮膚温を主とした単調作業ストレスの評価”, SICE第11回ヒューマン・インタフェース・シンポジウム論文集, pp.237-240(1995).
- 3) 長谷川浩史, 那須靖弘, 志水英二, “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”, 信学技報, PRU89-26, pp.57-62(1989).
- 4) D.J.Beymer, “Face Recognition Under Varying Pose”, A.I Memo, No.1461, Artificial Intelligence Laboratory, Massachusetts Institute of Technology(1993).
- 5) 宋 欣光, 李 七雨, 徐 剛, 辻 三郎, “部分特徴テンプレートとグローバル制約による顔器官特徴の抽出”, 信学論(D-II), vol.J77-D-II, No.8, pp.1601-1609(1994).
- 6) 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”, 信学技報, IE91-3, pp.17-24(1991).
- 7) 保黒政大, 長坂保典, 梅崎太造, 鈴木宣夫, “ニューラルネットワークによる顔の特徴点検出の評価”, 信学技報, PRU94-75, pp.23-30(1994).
- 8) Rumelhart D.E., McClelland J.L. and the PDP Research Group, “Parallel Distributed Processing”, 1, The MIT Press(1986).
- 9) R.O.Duda, P.E.Hart, “Pattern Classification and Scene Analysis”, Wiley, pp.271(1973).
- 10) 今村友彦, 藤吉弘亘, 梅崎太造, 金出武雄, “ニューラルネットワークによるナンバープレートの位置検出”, 信学技報, PRU95-68, pp.109-114(1995).