

# ニューラルネットワークによる回転を考慮した顔画像の識別

## Face Recognition Using Neural Network Under Varying Pose

○須藤覚\*, 大友照彦\*, 大槻恭士\*

○Satoru Suto\*, Teruhiko Ohtomo\*, Takashi Ohtsuki\*

\*山形大学

\*Yamagata University

キーワード : 顔画像識別 (face recognition)

局所結合型ニューラルネットワーク (locally connected neural network)

連絡先 : 〒992 山形県米沢市城南4-3-16 山形大学 工学部 共通講座 大友研究室

大友 照彦, Tel.: (0238)26-3371, Fax.: (0238)21-8193, E-mail: ohtomo@banana.yz.yamagata-u.ac.jp

## 1. はじめに

近年、金融機関の利用等におけるセキュリティシステムでは、種々の個人識別技術が利用されている。現在、最も普及している個人識別法は、パスワードやカードを使った識別である。この方法は簡単で実用的ではあるが、紛失や偽造という危険が伴う。そこで、数多くの個人識別技術<sup>1-3)</sup>の研究が行なわれている。

顔画像を用いた個人識別の利点は、本人の能動的行為を必要とせず、偽造が困難であるという点にある。この利点を生かすためには、人物をごく自然な形で識別できなければならない。しかし、顔を画像にした場合、視点の位置によって得られる画像は異なるものになってしまう。この問題を回避するためには、視点位置の影響を受けにくいシステムの構築が重要となる。

本研究では、顔の回転を考慮した顔画像識別を提案する。この識別法ではあらかじめ多くの視点から撮影された画像をいくつかのニューラルネッ

トワーク(以下NN)に学習させておく。次に先ほど学習させたNNに異なる二方向からの画像を入力する。その出力を二段目のNNに入力し、全体の学習を行なう。このように、一度各方向からの画像を学習しておくことによって、顔の回転の影響が小さい顔識別システムを構築できるものと考えられる。

予備実験として、二方向からの画像を入力する必要性と入力角度差についての検討を行なう。識別実験として、20人分の顔画像の識別を行ない、本手法の有効性を検証する。

## 2. 識別システムの概要

### 2.1 物体の回転に対応した識別システム

立体を識別する際、視点の位置によって見かけ上の画像が変化してしまうという問題が生じる。特に、顔画像は首の回転のために、常に同一ポーズでの画像を得ることは困難になる。

また、3次元の物体を2次元の画像に変換する

際、空間的な情報が失われてしまう。人間の神経系では、左右の目から別々の情報を取り入れることによって、3次元的情報を失うことなく外界の情報を得ることができる。このことから、3次元の物体を2次元の画像で表現するためには、2つの視点から物体をとらえることが必要であると言える。

物体の回転を考慮するためには、様々な角度から得られた画像のデータベースが必要となる。画像はデータの性質上かなりの記憶容量を必要とし、通常のデータベースを用いるのはあまり好ましくない。そこで、様々な角度から得られた画像をNNに学習させ、物体の持つ空間情報や形状などの情報をまとめておく。この結果、学習させたNNはデータベースの代わりを果たすことになる。

以上のことを考慮した学習モデルをFig.1に示す。また、学習はバックプロパゲーション法を用い、以下の手順で行なう。

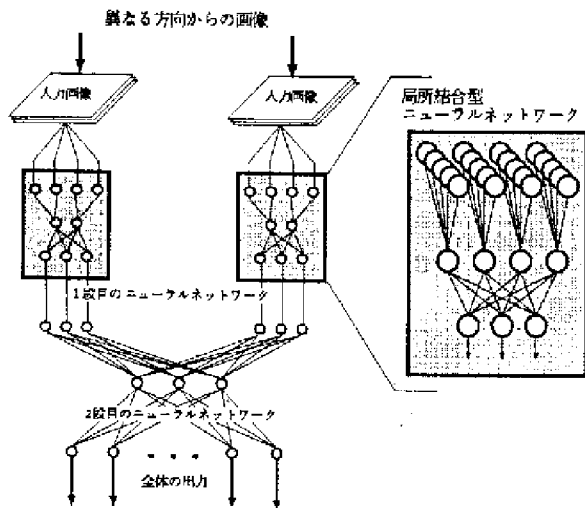


Fig. 1 物体の回転を考慮した識別システム

- 1) 様々な角度から得られた画像に対して前処理を行ない、1段目のNNに学習させる。
- 2) 1)のNNは2つ用意し、それぞれに異なる方向から得られた画像を入力する。
- 3) 2)の出力を2段目のNNの入力とし、2段目

のNNの学習を行なう。

## 2.2 全結合型NNと局所結合型NN

顔画像識別を行なうNNモデルは、3層の階層構造形で、層内結合を持たない層間全結合(Fig.2(左))が一般的である。全結合型NNによる顔画像の認識では、顔画像のデータが膨大となるため、NNの計算量が問題となる。また、顔画像の空間情報を処理するのに適した結合とはいえないため、良い認識率は得られないと考えられる。

本研究では、顔画像認識に適したNNを構築するために、局所結合型NN(Fig. 2(右))<sup>4)</sup>を採用する。全結合型NNは入力ユニットと各隠れユニットが全て結合しているため、このNNの隠れ層では画像の左上の情報も右下の情報も同一に扱われてしまう。これに対し、局所結合型NNは入力ユニットの一定領域が対応する隠れユニットとのみ結合している。すなわち、上位のユニットは下位の特定領域(受容野)だけから入力を得る。この結果、入力ユニット中の領域情報は他の領域情報と完全に分離され、画像空間的情報は保たれると考えられる。また、入力ユニットの一定領域に重なりを持たせることにより、多少の位置ずれを補正することができる。

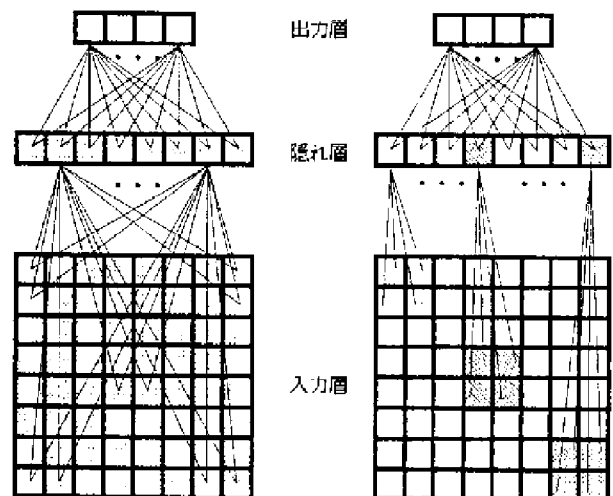


Fig. 2 全結合型NN(左)と局所結合型NN(右)

NNの構造を比較した場合、局所結合型NNの

結合数は全結合型 NN に比べて、非常に小さくなる。結合数の比較例を Table.1 に示す。この表からも分かるように、同じ数の隠れユニットを使った場合、結合数には明らかな開きがある。バックプロパゲーション法では、ユニット数と結合数に対して計算量が増加する。よって、局所結合型 NN を使った方が計算量は少なくなることが分かる。

	全結合型 NN	局所結合型 NN
入力画像	64 × 64	
受容野	64 × 64	8 × 8
隠れユニット	64	64
出力ユニット	6	6
結合数	262,528	4,480

Table 1 全結合型 NN と局所結合型 NN の結合数の比較

### 2.3 データ分割による学習

1 段目の NN はデータベースの役割を果たすことから、多くの画像を学習する必要がある。しかし、学習画像が多くなると学習に時間がかかるといった問題が生じる。また、学習画像を増やす場合にはもう一度始めから学習をしなければならない。これらの問題を回避するために、このシステムではデータをいくつかのグループに分割して学習するという方法をとる。

Fig.3 はデータ分割による学習の方法を表している。始めに、ある一定範囲の画像を 1 つのグループとし、それぞれのグループ毎に局所結合型 NN の学習を行なう。これを 1 段目の NN とする。1 段目の NN の学習が終わった後、異なる二方向から得られた画像を 1 段目の NN に入力し、その出力を 2 段目の NN に入力する。これを繰り返すことによって、2 段目の NN の学習を行なう。

この方法を用いた場合、並列化による学習時間の短縮がはかれる。また、学習画像の増加では、1 段目の NN を 1 つ追加することによって、簡単に対応することができる。

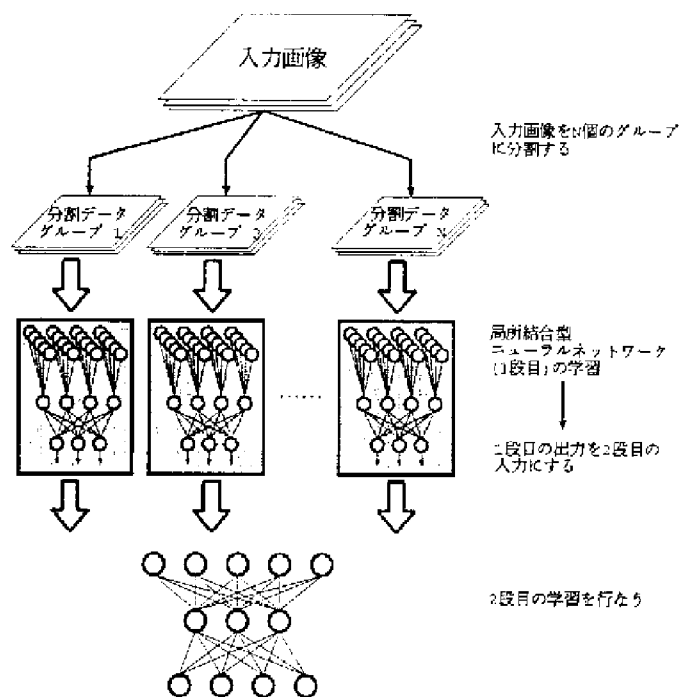


Fig. 3 データ分割による学習

## 3. データ採集と前処理

### 3.1 データ採集

本実験を行なうに当たり、20 人分の顔画像を Fig.4 の要領で撮影した。撮影対象の後ろには暗幕を張り、背景の影響をできる限り押える。回転椅子の下には角度を記した紙を置いておく。これにより、ほぼ同じ角度の顔画像が採集できると考えられる。また、照明条件を一定にするために自然光を遮断し、二方向から照明を当てるだけにする。

顔画像の撮影では、撮影対象に左手側を  $0^\circ$  とし、 $20^\circ$  から  $160^\circ$  まで  $10^\circ$  毎に回転してもらう。この時、各角度ごとに一点を注視してもらい、この状態をデジタルカメラで撮影した。

データは  $20^\circ$  から  $160^\circ$  まで  $10^\circ$  ごとに撮影したものを 1 セットとする。一人につき 10 セットの画像を撮影し、合計 3000 個のデータを採集した。なお、今回の撮影対象は、20 代前半の男性に限定した。これは性別や年齢の条件を一定にすることによって、なるべく同じカテゴリに属する顔画像を収集するためである。同一カテゴリに属する顔

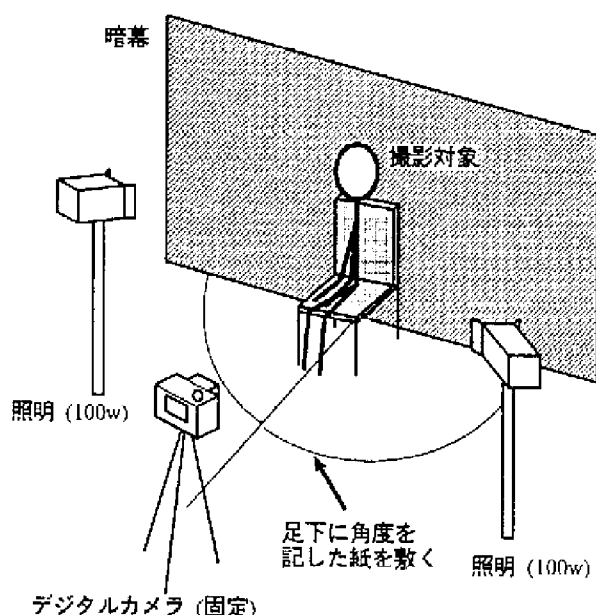


Fig. 4 顔画像の採集方法

画像の分類ができれば、性別や年齢などの他条件が加わることによって、分類に必要な情報量が増え、分類はより簡単に行なえるものと考えられる。

### 3.2 顔画像に対する前処理

3.1節で得られる画像は、 $420 \times 240$  のカラー画像である。この画像には顔以外の情報(髪、服等)が含まれているため、顔識別にそのまま使用するのは不適切と考えられる。また、顔の位置や大きさも一定ではないため、顔以外の要素による分類が行なわれる可能性もある。よって、顔以外の情報を除去し、顔の大きさをなるべく一定にする必要がある。

得られた画像は一つずつ手作業により、大雑把な顔の範囲を切り出し、 $128 \times 128$  の大きさにする。顔の範囲を切り出す時、なるべく髪部分は除く。なぜなら、髪型は一つの特徴と考えることもできるが、永続的な特徴ではないため、長期間の場合には誤認識の要因になると考えられるためである。

また、画像の認識にはカラー画像は必要なく、グレイスケールの画像で十分である<sup>1)</sup>という結果がある。したがって、カラー画像から256階長グ

レイスケール画像に変換し、データ量を減らす。

### 3.3 特徴抽出

3.2節での範囲切り出しの後、顔画像の特徴抽出を行なう。特徴抽出には、エッジ抽出とガウシアンを使用する。エッジ抽出では、顔の輪郭や大まかな目、鼻、口といった特徴を抽出するのが目的である。ガウシアンは顔の切り出し時に生じるわずかな位置ずれやスケールの変化を吸収するために、画像全体をぼやけさせるために使用する。以下に特徴抽出の手順を示す。

- 1) 顔画像のピクセル値の平均を求める。
- 2) 1)で求めた値を閾値とし、二値化を行なう。
- 3) 二値画像からエッジを検出する。
- 4) ガウシアンをかけ、画像全体をぼかす。
- 5) 認識用に  $64 \times 64$  の大きさにする。

3)の処理では、 $3 \times 3$  の大きさの局所オペレータを用いて、局所積和演算によりエッジを求める。エッジ検出オペレータは8方向 Laplacian を使用する。

4)のガウシアンは、式(1)に示す分散 $\sigma^2$ の2次元のガウス分布から求める。ここで、 $r$ は中心からの距離である。

$$G(r) = \frac{1}{2\pi\sigma^2} \exp\left(\frac{-r^2}{2\sigma^2}\right) \quad (1)$$



Fig. 5 特徴抽出の流れ

## 4. 顔識別実験

### 4.1 評価の定義

実験の評価として認識率、誤認識率、拒否率を用いる。1 段目の NN に入力する画像の組合せ数を Number、二段目の出力の中で拒否値以下のものを Reject、出力が拒否値を越え、出力ユニットが対応人物と一致するものを Success、出力が拒否値を越えるが、出力ユニットが対応人物と一致しないものを Error とした時、各評価を以下のように定義する。ここで、拒否値は出力の閾値であり、適当な値に設定しておく。

$$\text{認識率} = \frac{\text{Success}}{\text{Number} - \text{Reject}} \times 100 \quad (2)$$

$$\text{誤認識率} = \frac{\text{Error}}{\text{Number} - \text{Reject}} \times 100 \quad (3)$$

$$\text{拒否率} = \frac{\text{Reject}}{\text{Number}} \times 100 \quad (4)$$

### 4.2 実験手順

2.1,2.3節に示したモデルから実験・学習モデル (Fig.6) を作成する。このモデルを使用し、以下の手順により実験を行なう。ただし、ガウシアンのパラメータと拒否値は各実験によって変更する。

- 1) 濃淡画像 (128 × 128) から、エッジ画像を生成し、これにガウシアンをかける。
- 2) 1) で生成した画像を 64 × 64 に縮小し、入力画像を生成する。
- 3) 20° ~ 50°、60° ~ 90°、90° ~ 120°、130° ~ 160° のそれぞれの範囲について学習を行なう。学習は 10° 毎に行ない、この学習の結果を 2 段式 NN の 1 段目の結合係数として使用する。
- 4) 人数分の 20° 方向 ~ 160° 方向の画像をそれぞれの NN に入力し、学習を行なう。学習は 10° 毎に行ない、適当な角度差を持つ画像を組にして入力する。

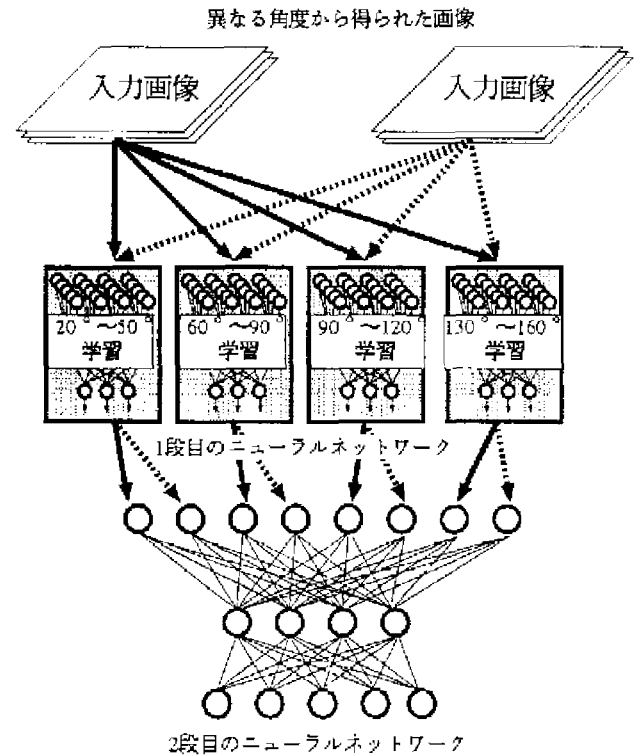


Fig. 6 実験・学習のモデル

- 5) 入力画像は一人につき 10 セット用意し、9 セットを学習用に、1 セットを識別用に使用する。これをそれぞれのセットについて繰り返し、10 回の学習・識別を行なう。
- 6) 学習から求められた結合係数を用いて識別モデルを構成し、未学習データを入力して識別実験を行なう。

### 4.3 入力画像数の検討

本システムでは入力画像を 2 枚使うことによって、より良い認識率を得られると考えている。しかし、入力画像がたとえ 1 枚でも良い認識率が得られる可能性もある。そこで、本研究で使用している 2 段式 NN と 1 段式 4 層階層型 NN の両方について実験を行ない、2 段式 NN の有効性を確かめる。

実験では 6 人分の顔画像を使用し、4.2 節の手順で行なう。ここで、ガウシアン分散を  $\sigma^2 = 3.0$ 、拒否値は 0.9 とする。また、各 NN のユニット数を Table.2 と Table.3 に示す。

1 段目		2 段目	
入力層(受容野)	10 × 10	入力層	48
入力層(重複部)	4 × 10		
隠れ層	100	隠れ層	30
出力層	6	出力層	6

Table 2 2 段式 NN のユニット数

入力層(受容野)	10 × 10
入力層(重複部)	4 × 10
第 1 隠れ層	100
第 2 隠れ層	20
出力層	6

Table 3 1 段式 4 層階層型 NN のユニット数

Table.4は、1 段式 4 層階層型 NN と 2 段式 NN の実験結果を示している。人数は少ないが、2 段式 NN では認識率 100% という良好な結果が得られた。また、出力の状況を見ても、同一人物だと判断した時の出力はかなり高く、他人物と判断した時の出力はほぼ 0 に近い。このことから、データの分類状況は良好だといえる。一方、1 段式 4 層階層型 NN は学習が完全にできていたにもかかわらず、未学習データを入力した場合に誤認識が生じた。拒否値を 0.9 と高めに設定しているため、分類が完全に間違っているものと考えられる。たとえば隠れ層の数を増やしたとしても、学習はこの状態でもできていることから、過学習になり認識率がより低くなる可能性の方が大きい。従って、回転を取り扱う場合には 1 方向の画像だけを使うよりも、2 方向から画像を取り込んだ方がより良い結果を得ることができるものと考えられる。

#### 4.4 入力角度差の変化に対する認識率

前節で 2 方向から画像を入力することの有効性が示された。次に、どの程度の角度差を持たせた時に良い識別ができるのかを検討する。入力角度差が小さい場合には、異なる方向からの画像を入

	4 層階層型 NN	2 段式 NN
認識率	699/735 = 95.1%	440/440 = 100.0%
誤認識率	36/735 = 4.9%	0/440 = 0%
拒否率	165/900 = 18.3%	100/540 = 18.5%

Table 4 1 段式 4 層階層型 NN と 2 段式 NN の認識率

力したとしても、入力画像の持つ情報はほぼ同じになる可能性もある。そこで、入力角度差を  $10^\circ$  ずつ大きくしていった時の認識率の変化を調べる。

実験では 6 人分の顔画像を使用し、4.2 節の手順で行なう。ここで、ガウシアン分散を  $\sigma^2 = 3.0$  とし、拒否値は 0 と 0.9 の 2 通りについて実験する。NN のユニット数は Table.2 と同じにする。

実験の結果を Fig.7 に示す。この結果から、入力角度差を大きくしていくにつれて、認識率が高くなっていることが分かる。拒否値を 0.9 にした場合、入力角度差が  $60^\circ$  以上で認識率が 100% になっている。つまり、2 方向から画像を入力する場合には、なるべく異なる画像を入力した方が良い識別ができることが分かる。拒否値を 0 にした時、この傾向はよりはっきり現れている。

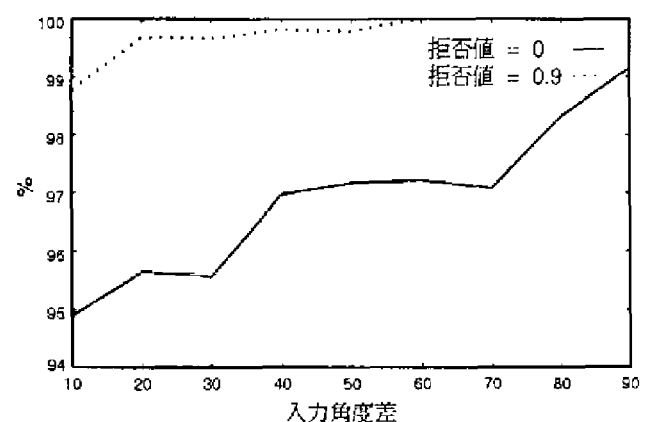


Fig. 7 入力角度差に対する認識率の変化

#### 4.5 人数増加時の識別実験

前節までの実験から、小人数の識別実験では 100% という良好な結果が得られた。ここでは人数

を増やした時の識別実験を行なう。

実験では20人分の顔画像を使用し、4.2節の手順で行なう。人数を増加することによって、識別率の低下が予想されるため、ガウシアン分散を $\sigma^2 = 7.0$ とする。また、1段目の出力層が増えるため、2段目の学習が困難になる可能性がある。よって、2段目の隠れ層をTable.5程度に増やすことにする。拒否値は0から0.9まで0.05ずつ増やし、識別率の変化をみる。

1 段目		2 段目	
入力層 (受容野)	10 × 10	入力層	160
入力層 (重複部)	4 × 10		
隠れ層	100	隠れ層	50
出力層	20	出力層	20

Table 5 4.5 節の実験における NN のユニット数

Fig.8に識別実験の結果を示す。拒否値を設けない場合でも90%以上の識別はできることから、良い分類が行なわれていることが分かる。しかし、拒否値を0.9と高めに設定した場合でも100%の認識率は得られなかった。これは人数の増加の影響もあるが、エッジ抽出の段階で似たような画像になったものがあつたためと考えられる。この点から、エッジ抽出の方法を改善するなどの対応が必要と考えられる。

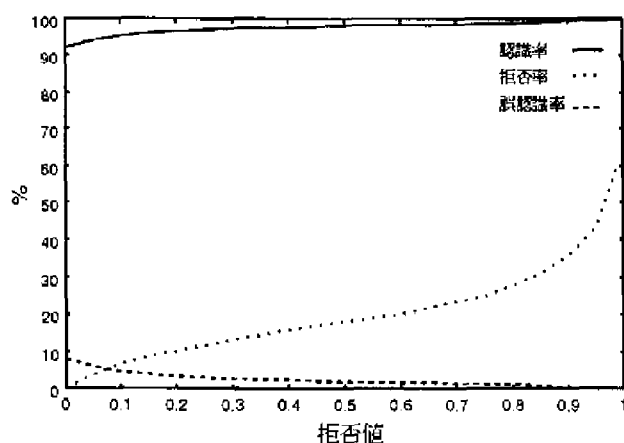


Fig. 8 顔識別実験

## 5. おわりに

本稿では、NNを用いて顔の回転に影響されにくい識別法を提案し、識別実験を通してその有効性を示した。

本方式では、顔の回転に強い識別を行なうために、1段目のNNに回転方向を考慮した学習をさせておき、識別時に異なる2方向からの画像を用いることによって、より精度の高い識別が可能になるというものである。また、学習モデルに局所結合型NNを用いることによって、学習時間を大幅に短縮させることができる。本方式を用いて20人分3000枚の顔識別実験を行なった結果、99.57%の識別結果が得られた。

本手法を用いることによって、回転に強い識別を行なえることが示された。今回の実験では、水平方向の回転のみを取り扱ったが、今後は垂直方向の回転を加えた場合について検討していきたい。

## 参考文献

- 1) Ashock Samal and Prasana A. Iyengar : "AUTOMATIC RECOGNITION AND ANALYSIS OF HUMAN FACES AND FACIAL EXPRESSIONS: A SURVEY", Pattern Recognition, Vol.25, No.1, pp.65-77, 1992.
- 2) David J. Beymer : "Face Recognition Under Varying Pose", Massachusetts Institute of Technology, C.B.C.L.Paper No.89, 1993.
- 3) 小杉 信 : "モザイクとニューラルネットを用いた顔画像の認識", 情報処理学会論文誌, vol.J76-D-II, no.6, pp.1132-1139, June 1993.
- 4) 大友照彦, 大槻恭七, 楊青, 石谷幹夫, 原健一 : "局所結合型神経回路網モデルによる手書き漢字の効率的認識法", 情報処理学会論文誌, Vol.35, No.6, pp.1091-1100, 1994.