

## ニューラルネットワークを利用した 濃淡画像からの衣服判別方式

河内谷 幸子

青山学院大学理工学部経営工学科  
〒157 東京都世田谷区千歳台6-16-1

あらまし 人物写真を濃淡画像として読込みニューラルネットワークを用いて衣服領域を抽出する方式について述べる。本システムは①入力画像の大きさ正規化を行う目的でいくつかの異なる大きさの顔領域教師信号を与えて顔の大きさを検出するネット、②濃淡画像により衣服領域を抽出するネット、③エッジ画像により衣服領域を抽出するネット、④これら②③の結果を統合するネット、の4種類のニューラルネットワークからなる。領域抽出に必要な最低限の階調と分解能を実験により求めた。入力データに条件をつければ本方式は良好な汎化能力を示す。和文キーワード：ニューラルネットワーク、濃淡画像、領域抽出、階調、分解能

## Recognition of Clothes from Shading Images Using Neural Networks

Sachiko KAWACHIYA

Faculty of Science and Engineering, Aoyama Gakuin University  
6-16-1 Chitosedai, Setagayaku, Tokyo 157, Japan

**Abstract** This paper describes the region extraction method from shading images. Our system has 4 neural networks. The first one detects the size of face region in order to regularize the size of input image. The second one extractes the clotes region by shading image. The third one extracts the clotes region by the edge image converted from the original image. The fourth one integrates the results from the second and the third ones. Experiments shows the minimum gray scale and quantization. This system recognizes the sample images correctly when those images follow some restrictions.

英文 key words : neural network, shading image, region extraction,  
gray scale, quantization

## 1. まえがき

縫製工場におけるデザイン業務の一手法として服飾雑誌中の写真を見てデザイン画を書く方法がある。この方法は2段階に分けられる。第1段階で写真を見て図1(a)のように人間が着衣したままの形状で絵画的な図を描き、第2段階で図1(b)のように一品ごとのデザイン図を作成する。図1(a)のような図をデザイン原画、図1(b)のような図をパーツ画とよぶことにする。パーツ画はバクンナーとよばれる型紙作成者によって型紙へと変換するため、重要な図である。

上記の2段階を計算機による画像処理におきかえると

- ・第1段階：エッジ抽出を中心とした処理
- ・第2段階：パーツ分離+後処理

となる。

本研究では写真からパーツ画を作成するための基本として写真を濃淡画像として読み込み、領域別のエッジ画像を得る手法について述べる。図1(b)のようなパーツ画を完全に得るためには本方式で得た領域別エッジ画像に対し、さらに何らかの手法を加えなければならない。例えば不要な線分の消去、見えない線分の追加等の

処理を画像の特徴抽出と知識ベースを組み合わせで行う手法が考えられる。本稿ではそのような後処理については考えず、領域分離という点に注目する。領域抽出に関しては集合写真の中から顔領域を抽出する研究<sup>1)</sup>にニューラルネットワークを用いて効果をあげたという報告<sup>2)</sup>がある。本稿もニューラルネットワークを利用して写真から衣服の領域を抽出することを考える。

ニューラルネットワークによる画像処理では入力画像の大きさや濃度について考慮しなければならない。Harmonによれば人間は8階調あればわずか16×16の正方形の集合から個人の顔識別ができるという<sup>3)</sup>。ニューラルネットワークを用いる場合にはどの程度の分解能と階調を必要とするであろうか。本稿では衣服領域抽出に必要な階調と分解能に関する興味深い結果を紹介する。また、ニューラルネットワークを利用する場合にはエッジ画像と濃淡画像の組み合わせ<sup>4)</sup>や分解能の細かいものと荒いものの組み合わせ<sup>5)</sup>のように異質画像による2種類の認識結果を組み合わせるとどちらか一方を単独で用いるよりも認識率が高まることが確認されている。本稿では異質画像の組み合わせについて領域分離に適したものを実験的に求めた。

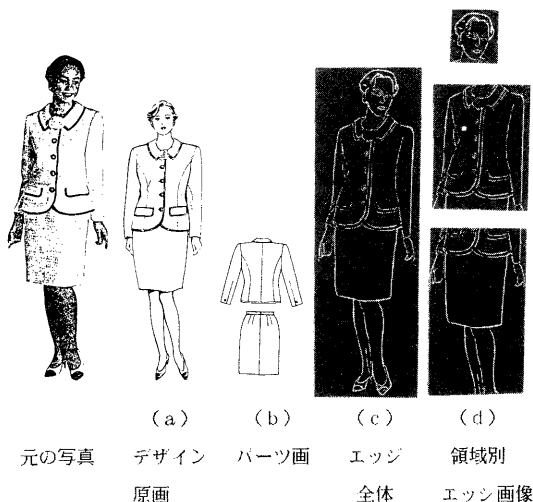


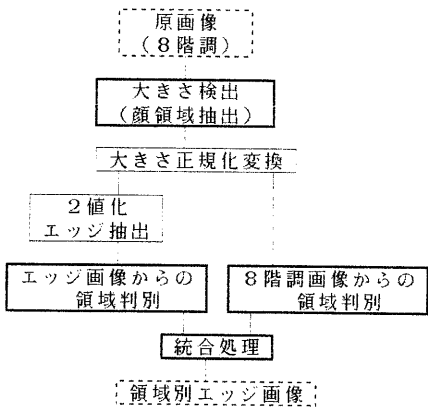
図1 デザイン画と領域別エッジ画像

## 2. システム構成

### 2.1 処理構成

本システムの処理構成を図1に示した。

[step1]原画像を8階調でスキャナから読み込む。  
[step2]大きさ検出のニューラルネットワークを用いて顔領域をもとに大きさをわりだす。本システムの特徴は領域検出において画像ごとに形状の変化が大きい領域と変化の少ない領域があることに注目し、変化の少ない領域だけを先に検出してこれを用いて大きさの正規化を行う点である。具体的には衣服判定の場合、衣服の形状は多種にわたるが、顔領域は個人差程度の変化である。しかも1冊の服飾誌において一般に衣服は50~200種類あるがそれを着用する人間(ファッションモデル)は5~7名である。3章で詳述するが、実験の結果から顔領域の認識率が他の領域に比べて高いことが確認できたので、顔領域を用いて大きさを検出する。このニューラルネットワークは入力層数が顔領域のウ






(注)  
 : ニューラルネットワークによる処理  
 : その他の処理  
 : 画像

図2 処理構成

インドウィの大きさ  $M \times N$  で出力層数が大きさの分類数  $k$  である。顔の大きさの異なる  $k$  種類の教師信号を与える。

[step3] 大きさ検出の結果から原画像の大きさを補正する。補正した画像は同じものを2画像用意する。これを画像  $a$  と画像  $b$  とする。

[step4] step3で作った画像  $a$  を2値化し、Prewittの方法(付録)でエッジ検出する。これを画像  $a'$  とする。

[step5] step4で作ったエッジ画像  $a'$  についてニューラルネットワークを用いて領域認識する。入力層数はウインドウの大きさ  $M' \times N'$  で出力層数が領域分類数  $k'$  である。形状の異なる  $k'$  種類の教師信号を与える。

[step6] step3で作った画像  $b$  についてもstep5と同様にニューラルネットワークで領域認識する。入力層数はウインドウの大きさ  $M' \times N'$  で出力層数は領域分類数  $k$  である。形状の異なる  $k'$  種類の教師信号を与える。

[step7] エッジ画像による領域認識結果と8階調画像による領域認識結果を統合ニューラルネットワークでまとめる。入力層数は

step5の出力層数  $k'$  + step6の出力層数  $k$  =  $2 \times k'$

である。また、出力層数は  $k'$  である。

以上により領域別のエッジ画像が得られる。

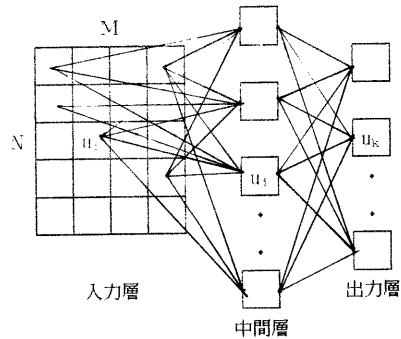


図3 ニューラルネットワーク模式図

## 2.2 ニューラルネットワークの構成と学習

本システムでは図3に示すように入力層、中間層、出力層の3層からなるニューラルネットワークを用いてBP法(誤差逆伝搬法)による学習を行った。以下にBP法の基本動作を述べる。

中間層  $u_j$  への入力  $U_j$  は、入力層  $u_i$  からの出力  $o_i$  に重み  $w_{ji}$  をかけたものである。学習効果を高めるためにさらにバイアス値  $b_j$  を加算する。

$$U_j = \sum w_{ji} o_i + b_j$$

同様に出力層  $u_k$  への入力  $U_k$  は、中間層  $u_j$  からの出力  $o_j$  に重み  $w_{kj}$  をかけて、バイアス値  $b_k$  を加えたものである。

$$U_k = \sum w_{kj} o_j + b_k$$

中間層の出力  $o_j$ 、出力層の出力  $o_k$  は

$$o_j = f(U_j)$$

$$o_k = f(U_k)$$

のように計算される。ここで  $f(x)$  はシグモイド関数を用いる。

$$f(x) = 1/(1+\exp(-x))$$

出力層  $u_k$  に与えられる教師信号を  $t_k$  とすると、1回の学習に対する重みの変化量は

$$\Delta w_{kj} = \alpha \delta_k o_j$$

$$\Delta w_{ji} = \alpha \sigma_j o_i$$

となる。 $\alpha$  は学習係数で  $\delta$  と  $\sigma$  は

$$\delta_k = f'(U_k)(t_k - o_k)$$

$$\sigma_j = f'(U_j) \sum \delta_k w_{kj}$$

である。

さらに学習効果を高めるためにn回目の重みの変化量を求める際に(n-1)回目の学習の重みを考慮する。ハイアス値も同様に考える。

$$\Delta w_{kj}^n = \alpha \delta_k^n o_j^{n-1} + \beta w_{kj}^{n-1}$$

$$\Delta w_{ji}^n = \alpha \sigma_j^n o_i^{n-1} + \beta w_{ji}^{n-1}$$

$$\Delta b_k^n = \alpha \delta_k^n + \beta b_k^{n-1}$$

$$\Delta b_j^n = \alpha \sigma_j^n + \beta b_j^{n-1}$$

$\alpha$ と $\alpha'$ は学習係数、 $\beta$ と $\beta'$ は慣性係数であり、その値は実験的に求める。4章に適用例を記述した。

### 3. 入力画像の階調と分解能

処理時間をできるだけ短縮するためには入力画素数(ウインドウの大きさ)および階調が少ない方がよい。いま上半身の領域を上方衣服領域、下半身の領域を下方衣服領域と定義する。これらの領域判定のために次のような実験を行った。

#### 実験1

まず大きさを正規化した画像を用意する。この画像について2階調、8階調、16階調の3階調と、2階調でスキャンした画像を既存の方法(付録)でエッジ抽出したもの、の4種類について考える。これら4画像を異なる3つの分解能で表現する。顔領域(Face Region)を4画素×4画素で表す分解能をF4、8画素×8画素で表す分解能をF8、16画素×16画素で表す分解能をF16とよぶことにする。顔領

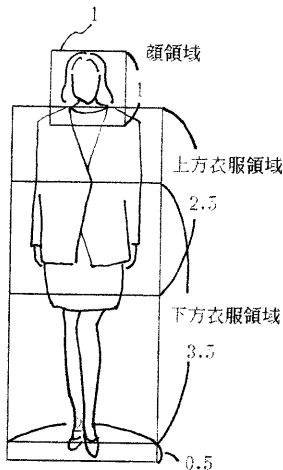


図4 各領域の大きさの比率

域以外の領域のウインドウはもっと大きくなる。識別したい上方衣服領域と下方衣服領域を具体的に図4のように定義する。

顔領域を1×1とすれば、図4に示すように上方衣服領域は2×2.5、下方衣服領域は2×3.5の大きさとなる。ウインドウの大きさは上方衣服領域の場合

F4: 8画素×10画素

F8: 16画素×20画素

F16: 32画素×40画素

下方衣服領域の場合

F4: 8画素×14画素

F8: 16画素×28画素

F16: 32画素×56画素

となる。

4種類の階調(エッジの場合を含む)と3種類の分解能を組み合わせた12種類について、60枚のサンプルデータにより実験を行った。実験に用いたニューラルネットワークとその学習方式は2.2節で述べた方式である。

#### 考察

表1は階調別・分解能別の領域認識率である。顔領域はほかの領域と比べて認識率が高い。これは衣服ほどダイナミックな変化が少ないためと考えられる。従って大きさの正規化には顔領域を用いるとよいことがわかる。

表1 階調別・分解能別の領域認識率

	2階調	8階調	16階調	エッジ
F4	66.7	80.0	83.3	70.0
F8	81.7	96.7	96.7	88.3
F16	83.3	96.7	96.7	90.0

	2階調	8階調	16階調	エッジ
F4	66.7	75.0	75.0	73.3
F8	70.0	83.3	88.3	75.0
F16	71.7	90.0	90.0	78.3

	2階調	8階調	16階調	エッジ
F4	63.3	71.7	71.7	68.3
F8	70.0	78.3	80.0	73.3
F16	70.0	80.0	80.0	75.0

顔領域では8階調でF8の場合良好な認識率となり、実験した範囲内ではそれ以上階調や分解能を上げてても認識率の著しい向上は見られない。顔領域による大きさ正規化は8階調・F8で行うことに決定する。

以下階調がa階調で分解能がFbである画像を、「a階調・Fb画像」と表記する。

実験1の12種類の階調・分解能では顔領域以外の部分の認識率が低い。階調や分解能の異なる画像を組み合わせると、認識率や学習回数に効果がある<sup>4)</sup>ことが知られている。表1に示した12種類から2種類を選ぶ組み合わせは12C2=66通りあるが、本システムにおいて最終的に必要とするエッジ画像を積極的に利用するという意味で、先ずエッジ画像とそれ以外の濃淡画像を組み合わせることを考える。それで良好な結果が出なければ、濃淡画像同士の組み合わせも実験することにする。

## 実験2

エッジ画像と濃淡画像を別のニューラルネットワークで認識し、それぞれの出力を統合ニューラルネットワークの入力として最終的な判断を行うようにする。画像は実験1と同じ60枚のサンプルを用いる。濃淡画像として

階調3種類(即ち2, 8, 16階調)

×分解能3種類(即ちF4, F8, F16)  
=9種類

を考える。エッジ画像の分解能は濃淡画像のそれと等しくする。

## 考察

表2は実験2の結果である。エッジ画像と上方衣服領域、下方衣服領域ともにエッジ画像+8階調・F8画像で良好な認識率となり、実験の範囲内ではそれ以上の階調・分解能では著しい認識率の向上は見られない。よって衣服領域の抽出にはエッジ画像と8階調・F8画像を組み合わせる方法が適している。

ここで良好な値を得られたので、濃淡画像同士を組み合わせる方法の実験を省略する。仮に濃淡画像同士の組み合わせ実験をおこなったとして、2階調・F4画像と16階調・F16画像の組み合わせがよいという結論が出た場合、①顔領域検出のための8階調・F8画像、②2階調・F4画像、③16階調・F16画像、④

最終的出力であるエッジ画像、の4画像が必要となってしまう。判定の前処理時間が増大してしまう。

表2 エッジ画像を加味した場合の階調別・分解能別の領域認識率

(a) 上方衣服領域 単位%

	エッジ+ 2階調	エッジ+ 8階調	エッジ+ 16階調	エッジ のみ
F4	78.3	86.7	86.7	73.3
F8	85.0	95.0	95.0	75.0
F16	86.7	95.0	95.0	78.3

(b) 下方衣服領域 単位%

	エッジ+ 2階調	エッジ+ 8階調	エッジ+ 16階調	エッジ のみ
F4	76.7	90.0	90.0	68.3
F8	81.7	93.3	93.3	73.3
F16	83.3	93.3	93.3	75.0

## 4. 適用例と評価

### 4.1 入力データの条件

本適用例の衣服領域判定の入力データは次のような制限を持つ。

- ・スキャナ入力時に目視により濃度調節を行う。
- ・紙面の大きさはB5判程度以下とする
- ・紙面内は人間1人だけであり、その他の部分(背景)は無地または風景のぼかしである
- ・人間の頭部から足先部まで全体が含まれる
- ・人間は正面を向いて立っている
- ・人間は肘を曲げていない
- ・人間は頭髪があり、脱帽状態である
- ・下記教師信号にある8種類の服装から極端にはずれていない、特に半袖は除外している点に注意
- ・人間はだいたい図4の比率となる

入力データの具体例およびその領域別エッジ抽出画像は図1に示した。

教師信号は図5に示すようにの8個を選んだ。まず上方衣服領域として

①長袖の長丈上着(ブレザー、コート)

②長袖の短丈上着(セーター)

を与え、下方衣服領域として

- ③ 短丈スカート
- ④ 長丈スカート
- ⑤ スポン
- ⑥ 短丈スカート一部分
- ⑦ 長丈スカート一部分
- ⑧ スポン一部分

を与える。⑥～⑧の一部分とは長丈の上方衣服により下方衣服の一部が隠れている状態をさす。

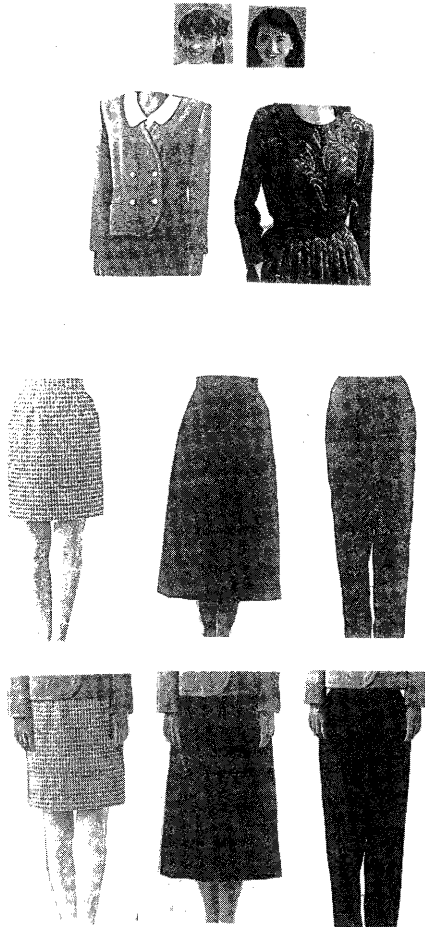


図5 教師信号

#### 4. 2 学習パラメータの設定

ニューラルネットワークの各パラメータは実験的に求めた。

#### 学習係数と慣性係数に関する実験

実験は3.1節で述べた顔領域の検出を8階調・F8で行う。まず中間層数はいくつかの例を参考に、十分な余裕のある値として30設けた。2.2節で述べた学習係数 $\alpha$ を0.1, 0.5, 0.9, 関数係数 $\beta$ を0.1, 0.5, 0.9として9通りの場合の実験を行い、収束状況を調べた。ここで $\alpha' = \alpha$ ,  $\beta' = \beta$ とした。その結果次のようなことがわかった。

- (1)  $\alpha$ が一定ならば $\beta$ は大きい方が収束が速い
- (2)  $\beta$ が一定の場合 $\alpha$ は大きい方が収束が速いが例外もある

最も収束状況の良かった組み合わせは学習係数 $\alpha = 0.5$ , 慣性係数 $\beta = 0.9$ である。

#### 中間層に関する実験

顔領域は8階調・F8, その他の領域はエッジと16階調・F16の濃淡画像で認識し、すべて学習係数 $\alpha = 0.5$ , 慣性係数 $\beta = 0.9$ とする。中間層数を網羅的に変化させて必要最低層数を求める。

以上の実験結果よりニューラルネットワークの各パラメータを表3のように定めた。

表3 ニューラルネットワークの各パラメータ

	大きさ 検出部	エッジ 利用部	8階調 利用部	統合部
入力層	8×8	16×28	16×28	16
中間層	20	25	25	20
出力層	3	8	8	8
学習係数	0.5	0.5	0.5	0.5
慣性係数	0.9	0.9	0.9	0.9

#### 4. 3 収束性

図6は表3のパラメータを用いた顔領域、上方衣服領域、下方衣服領域に対する収束曲線である。横軸は学習の反復回数である。縦軸は教師信号と出力信号の2乗誤差で、2.2節の定義を用いると誤差Eは、

$$E = (\sum (t_k - o_k)^2) / k$$

と表せる。kは出力層数である。

学習を10回反復することにEを計算した。このグラフからわかるように3領域とも大きな振動がなく、良好に学習が収束している。

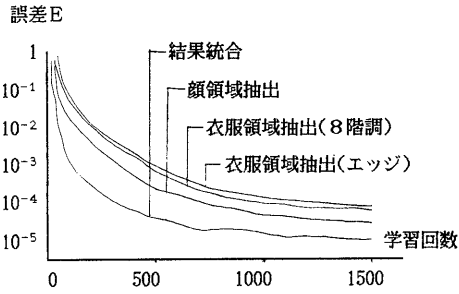


図6 収束曲線

#### 4. 4 顔領域による大きさ検出技法

大きさ検出ニューラルネットワークは入力層数が顔領域のウィンドウ画素数 $8 \times 8$ で出力層数が大きさの分類数3とした。サンプルデータを統計的に調査した結果、対象となる人間の大きさは、最小のものを縦横各4倍にしたものが最大のものとなることがわかった。ニューラルネットワークにおいては許容される大きさの変動はたかだか4倍である<sup>(6)</sup>ことを考慮して、第3章の実験1で用いた8階調・F8の画像と、これを縦横2倍に拡大したもの、4倍に拡大したもの、の3種類の教師信号を与える。ここでの縦横n倍の拡大は単純に注目画素と同一濃度で画素を置いていくものであるため3種類の教師信号は大きさは違っても分解能は等しい。

#### 4. 5 汎化能力

学習済みのニューラルネットワークが未学習の入力データに対しても期待するような出力を出すことを汎化能力という。これに関して次のような3つの実験を行った。実験におけるニューラルネットワークの諸環境は4.1節、4.2節で定めた通りである。

**汎化実験1**：人間が見て学習に用いたデータに近いと思う形状の未学習データについて認識実験を行う。ただしこの汎化実験1、2、3における「学習データに近い(又は異なる)」という判断基準は、実験を行った6名の判断により定めたものであり、データ数はそれぞれ20ずつである。

**汎化実験2**：人間が見て明らかに学習データと異なる形状の未学習データ、例えば和服や民族衣装、について認識実験を行う。

**汎化実験3**：学習データに近い服装であるが、人間が横向き・歩行中・拳手している等、姿勢の条件が4.1節で述べた条件に反するデータについて認識実験を行う。

#### 結果

汎化実験1に対しては学習データとほぼ同じ認識率を得られた。汎化実験2では60%前後の認識率であった。和服が認識できたのは興味深い。汎化実験3では認識率は極端に下がった。以上より4.1節で述べた条件の範囲内での汎化能力は良好だが、条件の範囲外での汎化能力は低いことがわかった。

#### 5. むすび

本稿ではニューラルネットワークを用いた濃淡画像の領域分離方式について報告した。入力画像に必要な最低限の階調と分解能を調べ、顔領域の抽出には8階調で8画素 $\times$ 8画素のウィンドウを用いれば充分であることがわかった。顔領域を8階調かつ16画素 $\times$ 16画素で表せば個人識別できる人間の能力に対し本研究は領域抽出なので問題が異なるものの、ニューラルネットワークの能力と人間能力との間に大差はないと考える。また、エッジ画像と濃淡画像を組み合わす場合に必要な濃淡画像の階調と分解能も求めた。これらの実験から、まず抽出率の高い顔領域を荒い分解能で抽出して大きさを求め、これを利用して画像の大きさ正規化を行い、顔以外の領域を前述の顔領域抽出処理より分解能の細かい濃淡画像およびエッジ画像の2画像から領域抽出を行う手法を提案した。入力データが一定の条件に従う場合の汎化能力は良好である。次の機会には中間層の動作分析について報告したい。

#### 謝辞

本研究の遂行にあたりご指導いただいた青山学院大学理工学部の実原実助教授、貴重なご意見とデータをくださったJUKI(株)縫製能率研究所第四課の浅井博康課長ならびに品川ファッションビジネス専門学校の嵯峨山温子校長に感謝する。

文献

- (1)安居院猛, 長尾智晴, 中嶋正之: “静止濃淡  
情景画像からの顔領域の抽出”, 信学論J74  
-DII, 11, pp1625-1627(1991).
- (2)安居院猛, 古久保良隆, 長橋宏: “神経回路  
網を用いた白黒集合写真からの顔領域の抽出  
”, 1991信学春季全大, 7, p268(1991).
- (3)吉川佐紀子(訳): “顔の認知と情報処理”,  
サイエンス社, pp59-64(1990).
- (4)佐野秀輝, 岩堀祐之, 石井直宏: “ニューラ  
ルネットワークによる濃淡画像識別システム  
の構築”, 信学技報, PRU91, pp9-16(1991).
- (5)河合秀夫, 田村進一: “ニューラルネットワ  
ークを用いた部分・モザイク顔による男女及  
び個人識別の一検討”, 1991情処秋期全大, 2,  
pp197-198(1991).
- (6)中野馨, 飯沼一元, 桐谷滋: “入門と実習  
ニューロコンピュータ”, pp237-238, 技術評  
論社(1989).
- (7)今井勝次, 郷原一寿, 内川嘉樹: “ニューラ  
ルネットワークの学習による目標パターンの  
位置, 大きさの検出”, 信学論J74-DII, 6,  
pp748-756(1991).
- (8)福島邦彦: “位置ずれに影響されないパター  
ン認識機能の神経回路モデル”, 信学論J62  
-A, 10, pp658-665(1991).

$$\begin{aligned}
 & - t4 \cdot f(i-1, j+1) \\
 & - t5 \cdot f(i, j+1) \\
 & - t6 \cdot f(i+1, j+1)
 \end{aligned}$$

ただし,  $s1, s2, s3, s4, s5, s6, t1, t2, t3, t4, t5, t6$  は重み係数である. これを加重マトリクスを用いて表すと図Aの(a), (b)のようになる. 図Aの(c), (d)のような重み係数を Prewitt オペレータという.

この方式も含めて一般に注目する画素の近傍領域を考慮する方法は雑音に強い.

\*は注目画素

s1		s4
s2	*	s5
s3		s6

(a)

t1	t2	t3
	*	
t4	t5	t6

(b)

1	0	-1
1	0	-1
1	0	-1

(c)

1	1	1
0	0	0
-1	-1	-1

(d)

図A 加重マトリクス

付録

[Prewittのエッジ抽出法]

いま座標(i, j)の濃度を  $f(i, j)$  で表すとする. x 方向, y 方向の1次の偏微分を  $\Delta_x f(i, j)$ ,  $\Delta_y f(i, j)$  とおく. デジタル画像処理の場合, 微分は差分で表現される. 注目する画素の近傍領域の濃度を考慮する方法として, 次のような計算方法がある.

$$\begin{aligned}
 \Delta_x f(i, j) = & s1 \cdot f(i-1, j-1) \\
 & + s2 \cdot f(i-1, j) \\
 & + s3 \cdot f(i-1, j+1) \\
 & - s4 \cdot f(i+1, j-1) \\
 & - s5 \cdot f(i+1, j) \\
 & - s6 \cdot f(i+1, j+1)
 \end{aligned}$$

$$\begin{aligned}
 \Delta_y f(i, j) = & t1 \cdot f(i-1, j-1) \\
 & + t2 \cdot f(i, j-1) \\
 & + t3 \cdot f(i+1, j-1)
 \end{aligned}$$