

特徴領域出力型ニューラルネットワークによる 郵便番号のセグメンテーション

横山 延雄*, 藤吉 弘亘**, 梅崎 太造**, 吉村 ミツ***

Segmentation of the Zip Code by the Artificial Neural Network which outputs a feature Area

Nobuo YOKOYAMA*, Hironobu FUJIYOSHI**, Taizo UMEZAKI**
and Mitsu YOSHIMURA***

(Received May 31, 1997)

In this study a method of segmenting the character area of the zip code using filter-net is proposed and the effectiveness is shown. Also, the inner structural characteristics of the network are investigated. In the conventional method, the processing of grazed characters, connected characters and the pre-processing of noise elimination, and so on, are necessary. However, these things are unnecessary in filter-net. The three filter-nets are made respectively for three-digit zip codes. In the filter-net, when segmenting the digit, the position and the size of another figure in addition to the area of the learning figure were learned and the estimation of the area of the learning figure was confirmed. In the evaluation experiment, when adding secondary searches, such as histogram processing to the output area of the filter-net, it was possible to segment the unknown data of 979 images by about 99 %.

Keywords : zip code, filter-net, segmentation, neural network

1. まえがき

郵便番号を高精度に認識するためには、まずその文字領域を正確に切り出すことが重要である。しかしながら、ワープロ等で印字される定型文字の郵便番号を除けば、一般的に郵便番号は、不特定多数の人により記入され、使用される筆記具もボールペンや筆等さまざまである。それゆえに、かすれ文字や続け文字、大きさの異なる文字が多く観測され、文字領域の切り出しを難しくしてい

る。

アメリカの郵便物には、決められた郵便番号の記入枠がないため、大きさや文字の間隔が異なり、一連の数字から1文字単位で文字を切り出すのは困難であるとされている^{[1][2]}。これに対して、日本の郵便物には予め郵便番号の記入枠があることから、枠周辺を探索するような切り出し方法を考えれば良い。通常は、濃度ヒストグラムのしきい値処理やラベリング処理等の簡単な切り出し法が用いられており、切り出し方法そのものの精度を十分に検討したものは少ない。また、かすれ部分が極端に多い郵便番号のデータは、評価用データから除外されている場合も少なくない。清水らは、かすれ文字処理や続け

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

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

***経営情報工学科 (Department of Management Information Engineering)

文字処理のプロセスを導入してセグメンテーション率を向上させているが, 明らかに1文字に見える場合でも, 論理的処理条件を満たさなければ切り出せないという問題がある^[3]。

近年, 従来の切り出し手法とは異なるニューラルネットワークを用いる方法が提案されている。今井らは, 複数のパターンを含む画像から, 目的領域を取り出す手法に, ニューラルネットワークを用いている。このネットワークは, 濃度の高い画素が集中している領域に反応するもので, 位置や大きさの異なる目的パターンの切り出しを行う際に有効な手段であり, フィルタネット(F.N)と呼ばれる^{[4]~[6]}。文字領域が長方形枠で出力されるため, かすれ文字処理や続け文字処理等が不要になる。フィルタネットの応用例としては, 藤吉らが, ナンバープレート上の数字を切り出す手法として用いており, 高い切り出し率を得ている^[7]。

本稿では, 数字領域を出力するように学習したフィルタネットを用いて手書き郵便番号の切り出しを行い, その有効性とネットワークの内部構造的特徴について報告する。

2. フィルタネットワーク

画像中に多くのノイズが含まれる場合, 従来より用いられているラベリング処理等で目標パターンのみを切り出すのは困難である^[8]。フィルタネットは, ノイズを含む画像や複数のパターンを持つ画像から, 目的パターンのみを抽出することができるという特徴を持つ。ここでは, 3桁の数字領域のみの切り出しに用いる。

2.1 フィルタネットの構成

使用するニューラルネットワークは, 3階層型で, 学習にはバックプロパゲーション法を用いる^[9]。ニューロンの人工的モデルであるユニットの応答関数には, 次式で表されるシグモイド関数を使用する。ここで, x はユニットへの入力を表す。 U_0 はシグモイド関数の傾きで, 大きいほどユニットの入力に対する反応は鈍い。

$$f(x) = \frac{1.0}{1.0 + \exp\left(\frac{-x}{U_0}\right)} \quad (1)$$

郵便番号の各桁に対応するネットワークを図1に示す。未学習データに対する汎化性と処理時間の高速化を考えて, フィルタネットの入力画像は1/10に圧縮している。そのため, 入力層のユニット数は288(24×12)となる。中間層のユニット数は, 64以上取れば学習誤差曲線が漸近的に減少関数的特性を示すが, ここでは余裕をみて128とする。出力層の数は入力層と同じ288とする。3つのフィルタネットは同じ構成である。

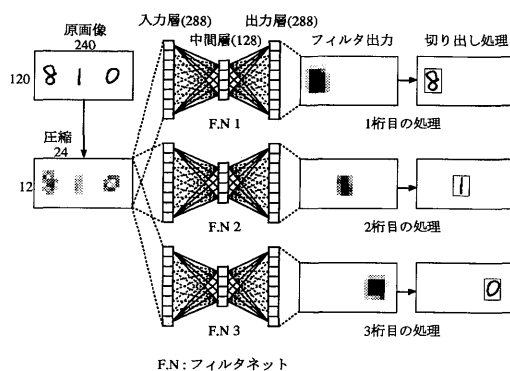


図1 ネットワークの構成

2.2 学習パターンと教師信号

学習パターンは, 各桁毎に調べた数字の出現頻度に基づいて20個選択する。視察により, かすれ文字や続け文字と判定されたデータは含まれていない。図2に示すように, 入力原画像中の文字データが含まれる範囲内で, 画像枠位置を毎回乱数によりずらして学習させるため, 基本的に同じパターンが学習に用いられることはない。外接長方形枠で切り出した数字領域を“1”(黒), それ以外を“0”(白)として教師信号を作成する(図3参照)。フィルタネットの出力が教師信号と等しくなるように, ニューラルネットワークの結合係数を修正する。予備実験の経験に基づき, 学習回数の打ち切りは十分誤差の収束する30000回とする。

2.3 学習誤差曲線

フィルタネットの学習を行う際の誤差曲線について述べる。フィルタネットの学習時, ネットワークの出力値と教師信号との差を誤差とする。 n 回学習したときの学習誤差 E_n は次式のように定義される。

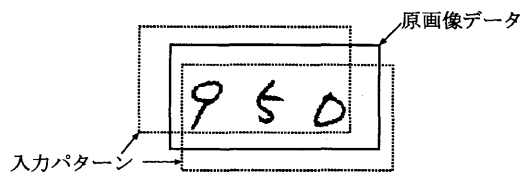


図2 学習パターンの作成

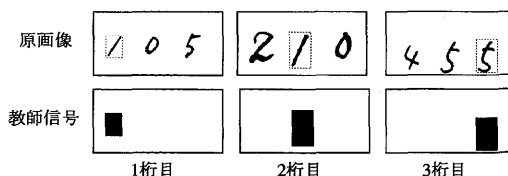


図3 学習パターンと教師信号

$$E_n = \frac{1}{N} \sum_{n=1}^N \sum_{i=1}^L \frac{(O_{n,i} - T_{n,i})^2}{2} \quad (2)$$

N : 学習データ数 (20)

L : 出力層のユニット数 (288)

$O_{n,i}$: 学習データ n が入力されたとき, 出力層のユニット i が出力する値

$T_{n,i}$: 学習データ n が入力されたとき, 出力層のユニット i に対する教師信号

2.4 評価方法

フィルタネットによる切り出し法の評価について述べる。図4に示すように, 教師信号 s_t と出力面積 s_o との排他的論理和を面積誤差 S_n として, 次式で定義する。

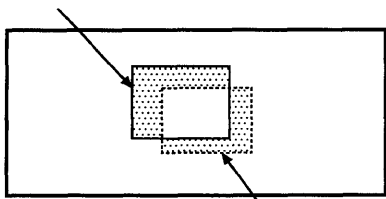
$$S_n = \frac{s_o \oplus s_t}{2} \quad (3)$$

フィルタネットの評価には, 評価データに対する S_n の平均値 $\overline{S}_n$ を用いる。

$$\overline{S}_n = \frac{1}{H} \sum_{n=1}^H S_n \quad (4)$$

ここで, H は評価用データ数である。 $\overline{S}_n$ が最も小さくなるように学習回数やシグモイド関数の傾き, 学習時のずらし幅等のパラメータを選択する。

目視により求めた領域 S_t



フィルタネットの出力領域 S_o

図4 面積誤差

2.5 評価基準

フィルタネットの切り出し率を算出する際の評価基準について述べる。郵便番号文字の大きさは一定ではなく, 単純に面積誤差 S_n の大小で切り出しの精度を評価することはできない。そこで, モが視察により求めた領域に対する相違度 ϵ を次式で定義する。

$$\epsilon_n = \frac{S_n}{S_t} \quad (5)$$

数字領域が75%以上含まれるとき ($\epsilon_n < 0.25$) を成功とし

て切り出し率を求める。なお, 切り出し成功と判断したデータの数字が読めるかどうか, 念のため視察により検査してある。

3. 画像データベース

3.1 画像データの仕様

画像データは, 郵政省郵政研究所所有の文字認識コンテストに用いられた公開サンプルデータで, 3桁の郵便番号からなる。横30[mm]×縦15[mm] (240×120[pixel]) のサイズで切り出されており, 予め2値化されている。郵便番号の枠は削除されている。画像サンプルを図5に示す。実験には, 続け文字やかすれ文字等を含む1000枚を使用する。学習用に各桁毎に20枚を使用する。評価用としては, 各桁毎に文字が含まれない画像1枚と学習データを除いた979枚を使用する。

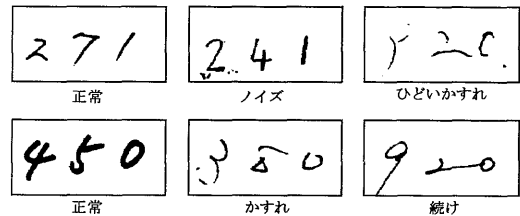


図5 郵便番号例

3.2 画像データの統計的性質

今回用いる郵便番号は, もともと枠内に記入されていることから, 一定の範囲内に分布している。故に, 分布領域を調べてフィルタネットを設計する際の有力なパラメータとして利用することが可能である。なお, 画像中の数字位置により, 桁番号を左から1桁目, 2桁目, 3桁目と定義する。

(1) 文字品質の内訳

各桁毎に調べた文字品質の内訳を表1に示す。かすれ文字が最も多く全画像枚数中5.2%を占める。

表1 品質の内訳 (画像 999 枚)

桁	ノイズ	かすれ	ひどいかすれ	続け文字
1桁目	9	47	25	0
2桁目	8	52	14	10
3桁目	6	57	14	10
合計 (%)	23 (0.8)	156 (5.2)	53 (1.8)	20 (0.7)

(2) 数字の出現頻度

各桁毎の数字出現頻度を表2に示す。1桁目と3桁目は, 2桁目と比較して各数字の出現頻度に偏りが生じている。

表2 郵便番号の出現頻度 (画像 999 枚)

桁	郵便番号の数字									
	0	1	2	3	4	5	6	7	8	9
1桁目	105	30	255	147	47	60	0	0	133	222
2桁目	26	211	69	84	96	56	166	70	215	6
3桁目	579	148	22	97	47	42	22	8	13	21

1桁目は“6”と“7”が全く含まれておらず, 3桁目は半数以上が“0”である。

(3) 郵便番号位置の分布

実験に使用する1000枚のデータを重ね合わせると図6ようになる。枠内に記入された郵便番号の位置分布は, ほぼ正規分布に従うことがわかる。

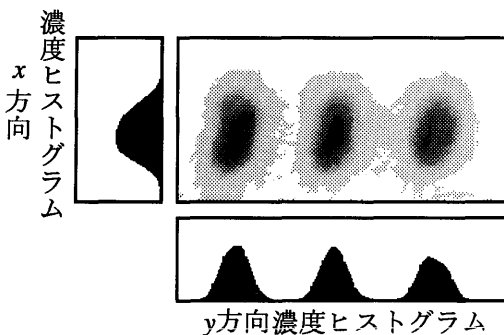


図6 郵便番号の出現頻度

(4) 重心座標の分布

画像データ1000枚分の郵便番号を視察により切り出す。各数字は外接長方形枠で囲まれるように切り出される。枠の重心を郵便番号の重心座標とする。2桁目の数字分布を図7に示す。各桁に対する重心座標の平均値($\bar{x}$, $\bar{y}$)と標準偏差 σ_x , σ_y を表3に示す。 σ_y は各桁ともほぼ近い値を示している。 σ_x は2桁目, 3桁目となるにつれ増加している。1桁目を特に注意して枠中に記入していることがその要因と考えられる。

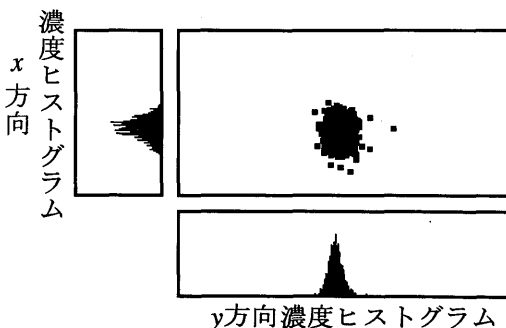


図7 2桁目数字の重心分布

表3 重心座標の平均値と標準偏差

桁	$\bar{x}$	σ_x	$\bar{y}$	σ_y
1桁目	40.5	4.9	71.5	7.4
2桁目	113.1	5.5	70.3	7.7
3桁目	185.9	6.1	69.6	7.6

4. フィルタネットの評価実験

4.1 学習範囲の検討

3.2節(4)で述べたように, 郵便番号各桁の分布を考慮してフィルタネットの学習を行う。乱数を用いて, 入力画像を任意の範囲内でずらしたパターンを作成し, フィルタネットに学習させる。ここで, ずらし幅は0, σ , 2 σ , 3 σ とする。各ずらし幅における数字の重心位置の学習範囲を図8に示す。ずらし幅を3 σ にすると, 図7で示される分布の99.3(992/999)%が含まれる。学習誤差曲線を図9に, 評価結果を図10に示す。ずらし幅が3 σ の場合, 誤差は減少しにくい, 未学習データに対しては面積誤差が一番少なく, 切り出し精度は高い。今回は割愛したが, 学習時のずらし幅を3 σ より大きくしてもほとんど変化は無いことが実験で確かめられている。以後の実験では, ずらし幅を3 σ とする。

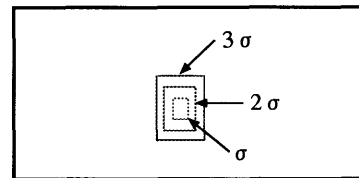


図8 学習範囲 (重心位置)

4.2 シグモイド関数の傾き

シグモイド関数の傾き U_0 を変えて実験を行う。図11に2桁目の学習誤差曲線を示す。 $U_0=0.2$ では神経の応答が鋭敏になり, 学習データの位置変動性に対して学習が追随できなくなる。 $U_0=1.0$ 以上では振動しながら減少して行くが, あまり U_0 を大きく設定しすぎると神経の応答が極端に鈍くなり学習が進まなくなる。図12に $U_0=1.0$ で学習したときの各桁に対する面積誤差を示す。3桁目を例にとると, 学習回数8000回辺りから過剰学習的な特徴が見られる。以後の実験での学習回数はそれぞれ2000, 13000, 8000とする。なお, U_0 は1.0とする。

4.3 切り出し結果

2.5節の評価基準により, 切り出し率を算出する。1桁目

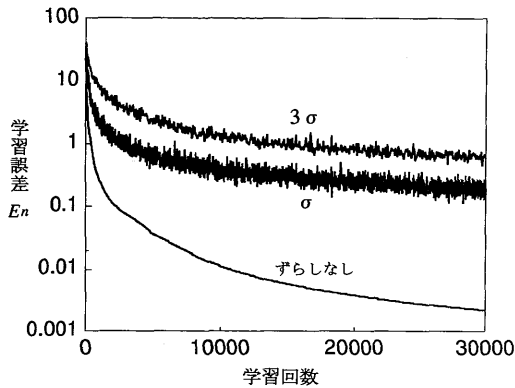


図9 学習誤差曲線 (2桁目)

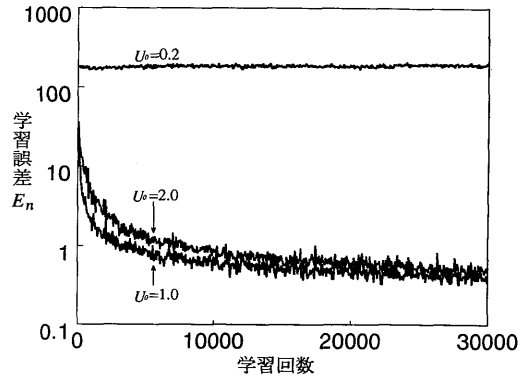


図11 学習誤差曲線 (2桁目)

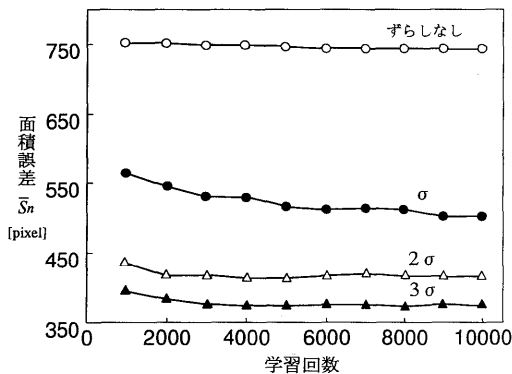


図10 面積誤差 (2桁目)

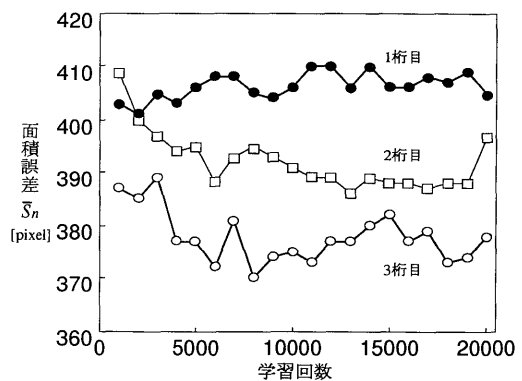


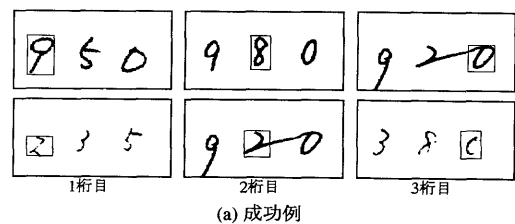
図12 面積誤差

の切り出し率が最も高く76.6% (750/979)である。2桁目70.8%, 3桁目63.7%と右に行くほど低下している。図13に切り出し成功例と失敗例を示す。なお、長方形の枠は切り出された領域である。

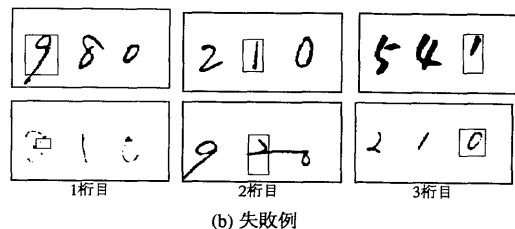
かすれ文字や続け文字は、特に学習させていないが、正確に切り出されている。失敗と判定された場合でも、視察により切り出した数字領域と比較して大きめに切り出されているものが多く、十分認識可能であると考えられる。一般に、手書き数字は縦長に書かれているため、上下方向に対しての分散値が大きく、切り出しを失敗している場合が多い。さらに切り出し精度を高めるためには、切り出し領域に対して再探索が必要である。

5. 荷重逆投影法

荷重逆投影の概念図を図14に示す。出力層ユニットの出力値を指定した後、式(6)に従い結合係数との積和演算を行い入力層ユニット I_k ($k=1,2,\dots,l$)への結合強度が計算される。ネットワークの荷重逆投影を行うことで、



(a) 成功例



(b) 失敗例

図13 フィルタネットの切り出し例

設計されたフィルタネットの特徴を視覚的に知ることができる^{[10][11]}。

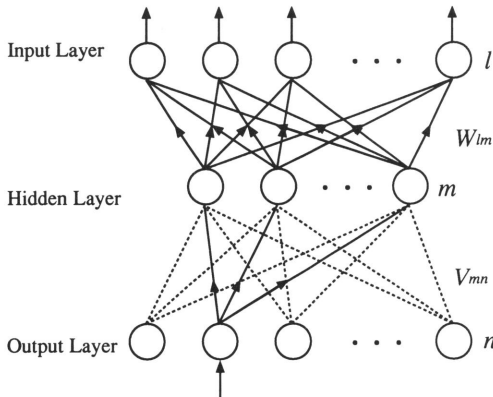


図14 荷重逆投影の概念

I_k : 入力層ユニット k への結合強度
($k=1,2,\dots,l$)

W_{lm} : 入力層—中間層間の結合係数

m : 中間層のユニット数

V_{mn} : 中間層—出力層間の結合係数

O_k : 出力層ユニット k の値 ($k=1,2,\dots,n$)

荷重逆投影の算出式を以下に示す。

$$\begin{bmatrix} I_1 \\ I_2 \\ \vdots \\ I_l \end{bmatrix} = \begin{bmatrix} W_{11} \cdots W_{1m} \\ W_{21} \cdots W_{2m} \\ \vdots \\ W_{l1} \cdots W_{lm} \end{bmatrix} \begin{bmatrix} V_{11} \cdots V_{1n} \\ V_{21} \cdots V_{2n} \\ \vdots \\ V_{m1} \cdots V_{mn} \end{bmatrix} \begin{bmatrix} O_1 \\ O_2 \\ \vdots \\ O_n \end{bmatrix} \quad (6)$$

5.1 学習回数に対する学習状態

出力層の1ユニットの出力だけを1として与えた場合の、学習回数に対する逆投影図を図15に示す。次式で示される I_k の最大値 λ により、正規化した値を濃度に変換して表示している。

$$\lambda = \max_{1 \leq k \leq l} I_k \quad (7)$$

学習回数の少ない初期段階では、学習効果が現れていないが、回数が進むにつれて2桁目に対する結合値は大きくなる。学習回数が増すと、2桁目のみでなく他の桁に対しても反応を示すようになる。500回あたりからは、濃度分布はほとんど変化しないが、 λ は少しずつ大きくなる。

5.2 フィルタネットの逆投影

各桁に対する逆投影例を図16に示す。図16-(a)より、出力が各桁の入力分布範囲内にある場合、ほぼ同位置に

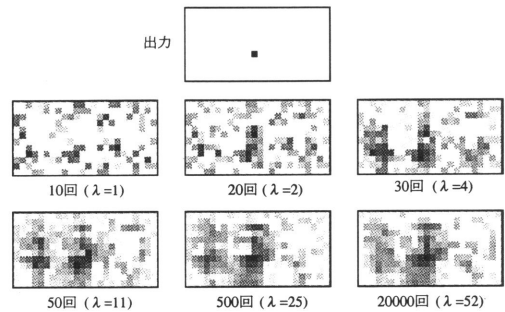
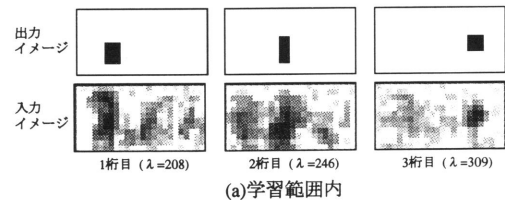


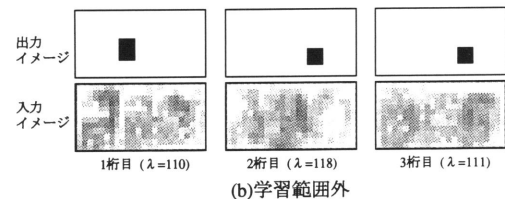
図15 学習状態

相当する部分への結合が最も強く現れているが、それ以外の桁にも反応していることがわかる。図16-(b)は、ニューラルネットワークに学習させた範囲外の領域に出力がある場合を想定したものである。 λ の値が(a)に比べて低く結合強度の高い領域が全体的に分散している。

図17は、学習させた2桁目のみが入力に含まれるパターンと、その逆のパターンがフィルタネットに通されたときの出力結果である。3桁すべてを入力したときの出力値に比べ、2桁目のみの場合、出力値は若干弱く、切り出し位置がずれている。学習桁以外の数字を入力した場合、2桁目の情報がないにもかかわらず出力が現れている。1桁目と3桁目の数字の位置と大きさから2桁目の数字領域を予測していることがわかる。



(a) 学習範囲内



(b) 学習範囲外

図16 逆投影の例

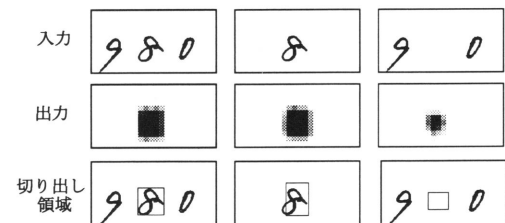


図17 フィルタネットの特性（2桁目を学習）

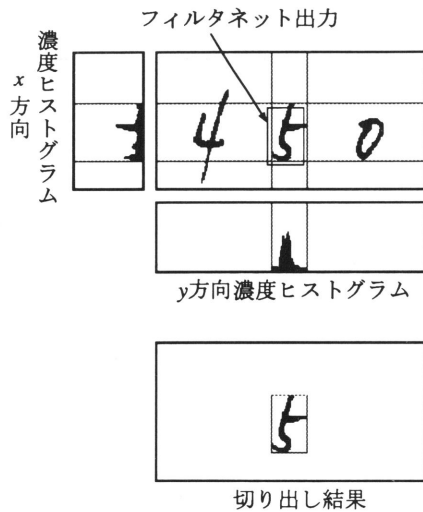


図18 二次探索例

6. 二次探索処理

6.1 探索方法

フィルタネットの出力に対して、更に切り出し精度を向上させるための二次探索処理を以下の手順で行う。

1. フィルタネットで切り出した枠の外に文字が続く場合、追跡処理を行い枠を拡大する。ただし、続け文字の存在を考慮して、左右端の追跡は5[pixel]以内とする。
2. 垂直・水平方向の各ライン毎に投影したヒストグラムにより、数字領域の上下左右端を決定する。

6.2 切り出し結果

視察により求めた領域に、さらに濃度ヒストグラムを用いて正確に切り出した領域を s_t 、二次探索後の切り出し領域を s_o として、2.5節で示す評価基準により比較した。1桁目の切り出し率は、98.4% (963/979)、2桁目99.6% (975/979)、3桁目99.1% (970/979) となり、フィルタネットの出力のみで切り出した場合と比較して約30%向上した(図19参照)。1桁目の面積誤差 $\overline{S}_n$ は、フィルタネットの出力のみで441[pixel]であるのに対して、二次探索後は20[pixel]となり、切り出し精度が大きく向上している。切り出し例を図20に示す。

切り出しに失敗した画像の多くはひどいかすれがある場合である。かすれ画像は濃度分布が低く、フィルタネットの出力が小さいため切り出しができない。

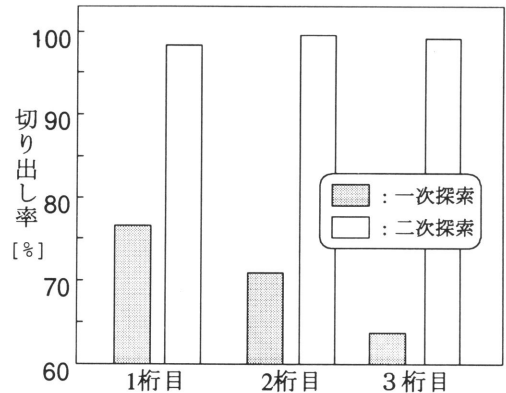


図19 一次及び二次探索能力

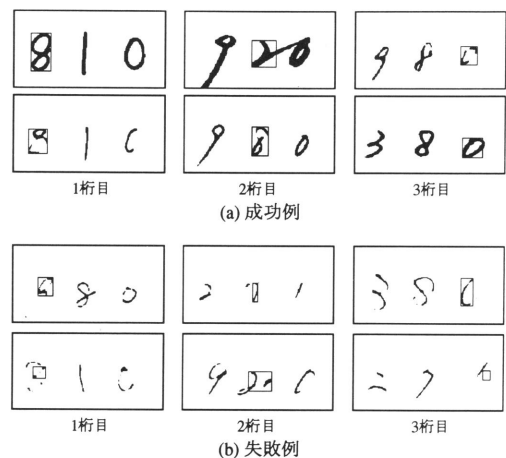


図20 二次探索後の切り出し例

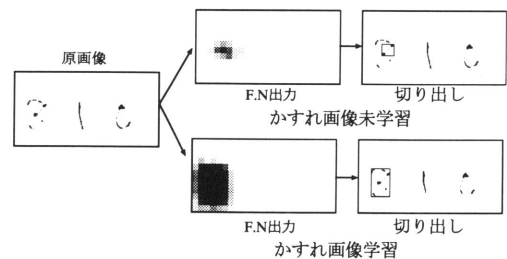


図21 かすれ文字を学習

上記の問題の対策として、最も切り出し率の悪い1桁目に対して、学習データにかすれ文字3枚を追加して学習させてみた。その結果、これまで失敗していた極端なかすれ文字の切り出しに成功するようになり、切り出し率99.8% (974/976)を得ることができた(図21参照)。

7. むすび

本稿では, 手書き郵便番号の切り出しにフィルタネットを用いて, その有効性を明らかにした。また, ネットワークの内部構造的特徴についても検討した。

まずはじめに, 学習時における教師信号のずらし幅とシグモイド関数の傾きに関して考慮して, 最適なネットワーク構造を考察した。1桁目の切り出し率が最も高く, 76.6% (750/979) を得た。

次に, フィルタネットの内部構造的特徴を荷重逆投影法により調べた。学習した桁の情報のみから出力しているのではなく, 残りの数字位置や大きさからも存在位置が推測されているのが確認された。

最後に, 濃度ヒストグラム等の二次探索処理を行うことで, 3桁とも約99%の高い切り出し率を得ることができた。さらに, かすれ文字を学習に追加することで, 極端なかすれ文字の切り出しが可能になり, 1桁目で99.8% (974/976) の切り出し率を得ている。

本手法を用いることにより, 従来手法では困難とされてきた“かすれ文字”や“続け文字”などに対しても簡単な処理で高精度な切り出しが行えることを明らかにした。

参 考 文 献

- [1] F.Kimura, Y.Miyake, and M.Shridhar: "Handwritten ZIP Code Recognition Using Lexicon Free Word Recognition Algorithm", ICDAR'95, pp906-910 (1995).
- [2] Yi Lu and M.Shridhar: "Character Segmentation in Handwritten Words-An Overview", Pattern Recognition, 1, pp.77-96 (1996).
- [3] 清水 達郎, 吉村ミツ, 佐々木一彦: "局所円弧パターン法による郵便番号読みとりシステム", 信学技報, PRU92 (1992).
- [4] 今井勝次, 郷原一寿, 内川嘉樹: "3層BPモデルを用いたノイズのある原画像に対する手書き文字認識", 信学論 (D-II), J75-D-II, 2, pp.314-323 (1992-02).
- [5] 今井勝次, 郷原一寿, 内川嘉樹: "3層BPモデルを用いた文字の切り出しと認識", 信学技報, PRU90-117 (1991).
- [6] 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と大きさの獲得", 信学論 (D-II), J74-D-II, 6, pp.748-756 (1991-06).
- [7] 藤吉弘亘, 梅崎太造, 金出武雄: "ナンバープレート内の一連番号の切り出しと認識について", 信学技報, PRU95 (1995).
- [8] 森 健一: "パターン認識", 第2章, 電子情報通信学会 (1988).
- [9] Rumelhart D.E., McClelland J.L. and the PDP Research Group: "Parallel Distributed Processing", 1, The MIT Press (1986).
- [10] 易 健強, 黒木秀一, 松岡清利: "誤差逆伝播学習法による位置ずれパターンの学習", 信学論 (D-II), J74-D-II, 1, pp.27-35 (1991-01).
- [11] 今井勝次, 郷原一寿, 内川嘉樹: "フィルタネットの内部表現について", 信学論 (D-II), J77-D-II, 2, pp.453-456 (1994-02).