

研究論文

多層競合ネットによる座標変換に不変なパターン認識*1

西田 健, 黒木 秀一

九州工業大学工学部*2

A Multi-Layered Competitive Net for Pattern Recognition Invariant to Coordinate Transformations

Takeshi Nishida and Shuichi Kurogi

Department of Control, Faculty of Engineering, Kyushu Institute of Technology

A multi-layered competitive neural network is presented for learning to achieve pattern recognition in a manner invariant to linear and/or nonlinear coordinate transformations such as projection, shift, rotation, magnification and so on. The transformed input patterns stored in the network are multiplied by the Jacobian of the transformation, an aspect shown to be essential for the transformation invariance. The network also has excellent generalization ability as has been verified by computer simulation.

1. はじめに

2次元平面パターンの様々な変形に対して不変な認識をニューラルネットを用いて実現しようとする代表的な研究として文献1)~5)などがある。これらの研究は、2次元画像理解のための包括的研究⁶⁾や実用的文字認識を目指す大規模ニューラルネットの研究⁷⁾と異なり、より基本的な不変認識の原理やアーキテクチャをその学習法とともに検討することに特徴がある。

これまでに提案されている不変認識を行うアーキテクチャとその機能には次のようなものがある。福島の新オコグニトロン¹⁾は視覚系を模倣した多層神経回路モデルを用いて位置ずれや局所的な変形に対して不変に認識することができる。Widrowらの多層ネット²⁾は、ADALINEと呼ばれる素子からなるインバリアンスネットと多層パーセプトロンからなり、平行移動、回転、拡大縮小や投影(水平方向のみの拡大縮小)に対する不変性を得るためのものである。GilesとMaxwell³⁾やPerantonisとLisboa⁴⁾のHONN(High-Order Neural Network)は、画素間の高次多項式で実現できる不変性を用いて平行移動、回転および拡大縮小に不変な認識を行う手法である。Konen

らのDLA(Dynamic Link Architecture)⁵⁾は多層ネットの層間の結合を動的に自己組織化することにより平行移動、回転、鏡面对称および歪みに対して不変認識を行う。一方、筆者らはこれらのネットとは異なるアーキテクチャをもつネットとして、アフィン変換に不変な認識を行う多層競合ネット⁸⁾を提案している。このネットは構造的にはWTA(Winner-Take-All)機能をもつ層を用いる点でネオコグニトロンに類似しており、クラスタと呼ばれるセルの集合ごとに同値なパターンを記憶させて不変性を実現するという点でslabと呼ばれるセルの集合を用いるWidrowらの多層ネットに類似している。しかし以上のネットは、いずれもアフィン変換に含まれる座標変換に対する不変認識を扱っており、非線形座標変換に対する不変認識を行うニューラルネットのアーキテクチャはこれまでに提案されていない。そこで本論文では上述の多層競合ネットを拡張し、非線形な座標変換群内の変換に不変な認識を行うことのできるネットを提案する。

ところで、非線形変換に対する不変認識の研究としてKanatani⁹⁾はカメラの回転変換(射影変換集合の部分集合の要素)に対する不変性について3次元シーンの画像復元の観点から研究を行っている。しかしそこで扱っているのは3次元シーンを構成する要素(点、線、テンソル、対称多項式、オプティカルフローなど)に対する不変性とその性質を用いた画像復元

*1 2000年5月11日受付, 9月7日受理

*2 〒804-8550 北九州市戸畑区仙水1-1

であり、本稿で扱うような画像パターン全体を一度に処理して不変認識を行う手法ではない。なお、本研究で不変性のために座標変換群を考えるのは、座標変換前後でパターンが属するクラスが同一でなければパターン分類が正しく定義できないからであるが、この問題設定は Kanatani⁹⁾ が指摘するように「幾何学的概念は座標変換群に関する不変性の性質の抽象化である」ことに関係しており、幾何学的パターンを扱う観点から一般性があると考えられる。

さて、不変認識を行う上述のすべてのニューラルネットは、その特徴的な学習機能として次のような汎化能力をもつ：まずいくつかの例を用いて不変性を学習させ、それを未知のパターンの不変認識に適用させることができる。さらに本論文で扱う多層競合ネット⁸⁾ は、ある変換のクラス（正確にはアフィン変換で構成した群）に関する不変性と識別能力を汎化学習する能力をもつことが示されている。より具体的には、この能力は TC 汎化 (Generalization of Training Classes) 能力と呼ばれ、つぎのように定義される：いくつかの訓練パターンをあるクラスに属するすべての変換により変形し学習させると、このクラスに属する変換のみに対する不変認識を未知のテストパターンに対して行うことができる。

本研究では、TC 汎化能力をもつ従来のネット⁸⁾ が適用可能な変換のクラスを一般の座標変換群にまで拡張し検討する。まず 2 では一般の座標変換に対する不変認識の方法を示し、非線形座標変換の例として射影変換について述べる。さらに、3 でネットの構造および 4 でその学習アルゴリズムを示す。最後に構成したネットの不変認識と TC 汎化能力についてのシミュレーション結果を 5 に示す。

2. 座標変換に対して不変なパターン認識

2次元空間 $\mathbf{x}=(x, y)^T \in \mathbb{R}^2$ 上でパターン $p_i(\mathbf{x})$ 、および座標変換 $\mathbf{g}(\mathbf{x})=(g_1(\mathbf{x}), g_2(\mathbf{x}))^T$ により変換したパターン $p_i(\mathbf{g}(\mathbf{x}))$ を考える。ここで $i=1, 2, \dots$ は区別すべきパターンの番号、 $(x, y)^T$ は行ベクトル (x, y) の転置すなわち列ベクトルを表す。またここで考えるパターンは文字などの濃度や輝度のパターンであり、 $p_i(\mathbf{g}(\mathbf{x}))$ は非負かつ有限であり、あるパターン $p_i(\mathbf{g}(\mathbf{x}))$ とその定数倍のパターン $ap_i(\mathbf{g}(\mathbf{x}))$ ($a>0$) は区別しない。さらに $p_i(\mathbf{g}(\mathbf{x})) \neq 0$ となるすべての点 \mathbf{x} は変換前後で、ある与えられた正方領域 $D = \{(x, y)^T \mid |x| \leq \Delta_b, |y| \leq \Delta_b\}$ の中にあるとする。

以下に、従来の線形変換に対する不変認識の手法⁸⁾

を拡張し、任意の座標変換群 G 内の変換 \mathbf{g} に対して不変認識を行うための手法を示す。ここで、線形変換を $\mathbf{g}(\mathbf{x})=A\mathbf{x}$ 、アフィン変換を $\mathbf{g}(\mathbf{x})=A\mathbf{x}+\mathbf{b}$ ($A \in \mathbb{R}^{2 \times 2}$, $\mathbf{b} \in \mathbb{R}^2$) として区別し解析する場合があるが本論文ではこれらの用語を区別せずに用いる。なお、以下では恒等変換を $\mathbf{e}(\mathbf{x})=\mathbf{x}$ とし、簡単のため $p_i(\mathbf{x})$ や $\mathbf{g}(\mathbf{x})$ などを p_i や \mathbf{g} などとも書く。

2.1 一般の座標変換に対する不変認識

まず座標変換 $\mathbf{g}(\mathbf{x}) \in G$ の Jacobian

$$|\mathbf{g}_x| = \left| \frac{\partial \mathbf{g}}{\partial \mathbf{x}} \right| = \frac{\partial g_1}{\partial x} \frac{\partial g_2}{\partial y} - \frac{\partial g_1}{\partial y} \frac{\partial g_2}{\partial x} \quad (1)$$

を訓練パターン $p_i(\mathbf{g}(\mathbf{x}))$ に乗じたパターン

$$p_{i\mathbf{g}}(\mathbf{x}) = p_i(\mathbf{g}(\mathbf{x})) |\mathbf{g}_x| \quad (2)$$

をすべての $\mathbf{g} \in G$ について作成し、それらを正規化したパターンの集合 $P_{iG} = \{p_{i\mathbf{g}} / \|p_{i\mathbf{g}}\| \mid \mathbf{g} \in G\}$ を記憶する。ここで $\|p(\mathbf{x})\| = (\int p(\mathbf{x})^2 d\mathbf{x})^{1/2}$ はパターン $p(\mathbf{x})$ のノルムを表す。また訓練パターン $p_i(\mathbf{x})$ の番号 i の集合を $I = \{1, 2, \dots\}$ とし、その要素数 $|I|$ は有限であるとする。さらに、パターンの非零値が存在する正方領域 D 内で $|\mathbf{g}_x|$ は非零、有限かつ連続であるとする。ここで座標変換 $\mathbf{z}=\mathbf{g}(\mathbf{x})$ が面積 S_x の微小領域 D_x を面積 S_z の微小領域 D_z に写すとすると

$$S_z = \int_{D_z} dz = \int_{D_x} |\mathbf{g}_x| d\mathbf{x} \approx |\mathbf{g}_x| S_x \quad (3)$$

より

$$|\mathbf{g}_x| \approx S_z / S_x \quad (4)$$

となることを注意しておく。これにより(2)式の記憶パターン $p_{i\mathbf{g}}(\mathbf{x})$ は座標変換により拡大(縮小)された部分のパターンの値を小さく(大きく)するように変換パターン $p_i(\mathbf{g}(\mathbf{x}))$ を修正していることが分かる。

次に、テストパターン $p_j(\mathbf{x})$ を $a(>0)$ 倍し、座標変換 $\mathbf{h}(\mathbf{x}) \in G$ を施した $ap_j(\mathbf{h}(\mathbf{x}))$ と $p_{i\mathbf{g}} / \|p_{i\mathbf{g}}\| \in P_{iG}$ との内積の最大値

$$S_{j(a, \mathbf{h})i} = \max_{\mathbf{g} \in G} \left\{ \int ap_j(\mathbf{h}(\mathbf{x})) \frac{p_i(\mathbf{g}(\mathbf{x})) |\mathbf{g}_x|}{\|p_i(\mathbf{g}(\mathbf{x}))\| |\mathbf{g}_x|} d\mathbf{x} \right\} \quad (5)$$

を計算する。ここで $j \in J$ であり、 J はテストパターン $p_j(\mathbf{x})$ の番号 j の集合である。すると $\mathbf{h}(\mathbf{x}) \in G$ であるから G の中には $\mathbf{z}=\mathbf{h}(\mathbf{x})$ の逆変換が存在するので、 $\mathbf{g}(\mathbf{x})=\mathbf{g}(\mathbf{h}^{-1}(\mathbf{z}))=\mathbf{f}(\mathbf{z})$ および $|\mathbf{f}_z| = |\partial \mathbf{g} / \partial \mathbf{x}| \cdot |\partial \mathbf{x} / \partial \mathbf{z}| = |\mathbf{g}_x| \cdot |\mathbf{h}_x|^{-1}$ を用いると、

$$S_{j(a, \mathbf{h})i} = \max_{\mathbf{f} \in G} \left\{ \frac{a}{\langle |\mathbf{h}_x| \rangle_i^{1/2}} \int p_j(\mathbf{z}) \frac{p_{i\mathbf{f}}(\mathbf{z})}{\|p_{i\mathbf{f}}(\mathbf{z})\|} dz \right\} \quad (6)$$

となる。ここで

$$\langle |\mathbf{h}_x| \rangle_i = \int \left(\frac{p_{i\mathbf{f}}(\mathbf{z})}{\|p_{i\mathbf{f}}(\mathbf{z})\|} \right)^2 |\mathbf{h}_x| dz \quad (7)$$

である。以下、この値がいくつかの仮定のもとで i に対してほぼ一定となることを示す。まず、(6)式の右辺の最大値を与える変換 $f \in G$ を f^* とし、パターン $p_{ir^*}(z)$ の非零領域を $D_{ir^*} = \{z | p_{ir^*}(z) \neq 0\}$ 、パターン $p_j(z)$ の非零領域を $D_j = \{z | p_j(z) \neq 0\}$ とする。さらに、通常の文字認識において仮定される平行移動や拡大縮小などを含む合成変換が G に含まれるとすると、(6)式の最大値を求める操作により D_j と D_{ir^*} はかなり大きな共通部分をもつことになる。このとき変換 $h(x)$ が線形 $z = h(x) = Ax + b$ ならば、 $\langle |h_x| \rangle_i = |A|$ は i に無関係な定数になる。変換 $h(x)$ が非線形ならば、 $\int (p_{ir}(z) / \|p_{ir}(z)\|)^2 dz = 1$ であるから、(7)式より $\langle |h_x| \rangle_i$ は D_{ir^*} における $|h_x|$ の最小値と最大値の中間値をとる。よって D_{ir^*} と D_j が大きな共通部分をもちかつ変換 $h \in G$ が連続であるという条件のもとでは、 $\langle |h_x| \rangle_i$ は D_j の中心付近における $|h_x|$ の値に近い値をもち i に対するばらつきはかなり小さくとなると考えられる。

そこで i に無関係な定数 a_h を用いて $\langle |h_x| \rangle_i^{1/2} \approx a_h$ とおくと、(6)式は

$$S_{j(a,h)i} \approx \left(\frac{a}{a_h}\right) \max_{f \in G} \left\{ \int p_j(z) \frac{p_{ir}(z)}{\|p_{ir}(z)\|} dz \right\} \\ = \left(\frac{a}{a_h}\right) S_{ji} \quad (8)$$

となる。ここで S_{ji} は変換前のテストパターン $p_j(z)$ と $p_{ir}/\|p_{ir}\| \in P_{iG}$ との内積の最大値、すなわち $h(x)$ として恒等変換 $e(x) = x$ を用いたときの $S_{j(1,e)i}$ である。すると $S_{j(a,h)i}$ を $i \in I$ について並べたベクトル

$$\mathbf{S}_{j(a,h)} = (S_{j(a,h)1}, S_{j(a,h)2}, \dots, S_{j(a,h)I}) \quad (9)$$

を正規化したものは、

$$\frac{\mathbf{S}_{j(a,h)}}{\|\mathbf{S}_{j(a,h)}\|} \approx \frac{(S_{j1}, S_{j2}, \dots, S_{jI})}{\|(S_{j1}, S_{j2}, \dots, S_{jI})\|} = \frac{\mathbf{S}_j}{\|\mathbf{S}_j\|} \quad (10)$$

となり、 a および h に関して不変な値をもつ。

さて、このベクトルを用いて不変認識を行うには不変性だけでなく異なるパターンを識別する能力も必要である。このための前提条件として、あるパターン番号 $k \in I \cup J$ をもつパターンが G に属する座標変換およびパターン値の定数倍により他のパターン番号をもつパターンになることはないとする。これは $k, l \in I \cup J$ に対して

$$\max_{g, h \in G} \left\{ \int \frac{p_k(g(x))}{\|p_k(g(x))\|} \frac{p_l(h(x))}{\|p_l(h(x))\|} dx \right\} \leq 1 \quad (11)$$

の等号が成立するのは $k=l$ のときに限ることを意味する。すると、以下のように、テストパターンに非常に良く似たパターンを記憶することにより識別能力が得られる。すなわち、極端な場合、 $I=J$ としてテストパターンをすべて記憶すれば、(6)式の積分の最大

値は(11)式の条件から、 $i=j$ および $f(z) = e(z) = z$ のとき得られ、その値は $S_{j(a,e)j} = a \|p_j(z)\|$ になる。従ってテストパターン $ap_j(h(x))$ に対する(10)式のベクトルはその第 j 要素のみが最大値をとるので、このパターンを識別できることになる。

以上は訓練パターンに内在する変換群 G のみに対する不変認識、すなわち TC 汎化による不変認識を行うための本手法の基本原則であり、第 j パターンを認識するには $\mathbf{s}_j / \|\mathbf{s}_j\|$ だけを記憶しておけばよいことが分かる。これにより各パターンが正しく不変認識されるかどうかは(8)式における不変性のための近似と(11)式左辺のパターン間の類似度との関係に依存する。

一方、実際の文字認識などでは(8)式により十分な不変性が得られないような変形文字や異なる書体文字を用いる場合もある。そのような場合にはそれらのパターン $ap_k(h(x))$ に対する $\mathbf{s}_{k(a,h)} / \|\mathbf{s}_{k(a,h)}\|$ も第 j パターンとして記憶することが必要となる。すなわち第 j パターンとして認識すべきパターンでありながら $\mathbf{s}_j / \|\mathbf{s}_j\|$ を用いて認識できない場合は、それらに対する $\mathbf{s}_{k(a,h)} / \|\mathbf{s}_{k(a,h)}\|$ も記憶する(追加学習する)ことにより所望の不変認識が可能になると考えられる。

2.2 射影変換とその Jacobian

射影変換 $g(x)$ は一般に次の式で表される非線形座標変換である¹²⁾。

$$g(x) = \frac{Ax + b}{p^T x + 1} \quad (12)$$

ここで、

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad b = \begin{pmatrix} e \\ f \end{pmatrix}, \quad p = \begin{pmatrix} p \\ q \end{pmatrix} \quad (13)$$

であり、 $a, b, c, d, e, f, p, q \in \mathbb{R}$ である。なお $p=q=0$ とすると射影変換は線形変換となることが分かる。また射影変換されたパターン $p_i(g(x)) = p_i(z)$ は射影変換により変換前の点 z が変換後に $x = (zp^T - A)^{-1}(-z + b)$ に写る。例えば、(13)式の各パラメタが $a=1, d=1, p=0.03$ でその他が零のとき、Fig. 1(a)のパターンは同図(b)のパターンに射影される。 $p_i(z) = p_i(g(x))$ により点 $z = (x', y')^T$ は点 $x = g^{-1}(z) = (x', y')^T / (1 - 0.03x')$ に写る。なおこの例および後述するシミュレーションでは Fig. 1(a)と同程度の大きさのパターンが正方領域 $D = \{(x, y)^T | |x| \leq 30, |y| \leq 30\}$ 内に存在するように $p=0.03$ 等のパラメタの上限値が決定されたことを注意しておく。

次に、(12)式で与えられる射影変換 $g(x)$ の Jacobian は、

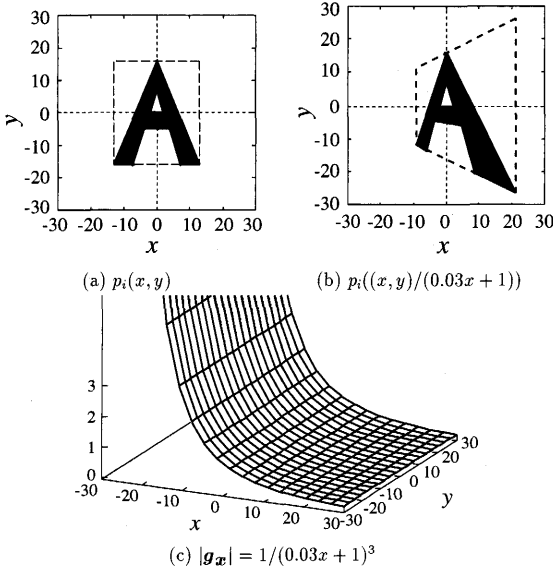


Fig. 1 Example of the relation between a projective transformation and the Jacobian. The pattern values at white and dark areas of (a) and (b) are zero and non-zero constant, respectively.

$$|g_x| = \frac{|M|}{(\mathbf{p}^T \mathbf{x} + 1)^3} \quad (14)$$

となる。ここで M は射影変換行列

$$M = \begin{pmatrix} A & \mathbf{b} \\ \mathbf{p}^T & 1 \end{pmatrix} \quad (15)$$

である。従って、例えば前述の例で与えられる射影変換に対する Jacobian は Fig. 1(c) のようになる。

3. ネットワーク

ここで用いるネットはアフィン不変認識を行う多層競合ネット^{8,10,11)}と同じく Fig. 2 に示す 3 層構造をもつ (ただし Fig. 2 は認識時の様子を表し、学習時の信号はこれとは異なることに注意)。第 2 層と第 3 層は複数のクラスタから成り、各クラスタは複数の競合セル (competitive cell) と 1 つの統合セル (integration cell) をもつ。

まず前節で考えた連続空間 \mathbb{R}^2 のかわりに $N \times N$ 画素の離散空間を考える。すなわち実際のネットは有限の容量しかもたないため、第 1 層においては \mathbb{R}^2 空間における入力パターンを N^2 次元ベクトルに離散化して出力する。ここで訓練パターン $p_i(\mathbf{g}(\mathbf{x}))$ の学習時には (2) 式を離散化した

$$\mathbf{p}_{ig}^{(2)} = (\mathbf{p}_{ig}(x_1), \mathbf{p}_{ig}(x_2), \dots, \mathbf{p}_{ig}(x_{N^2}))^T \quad (16)$$

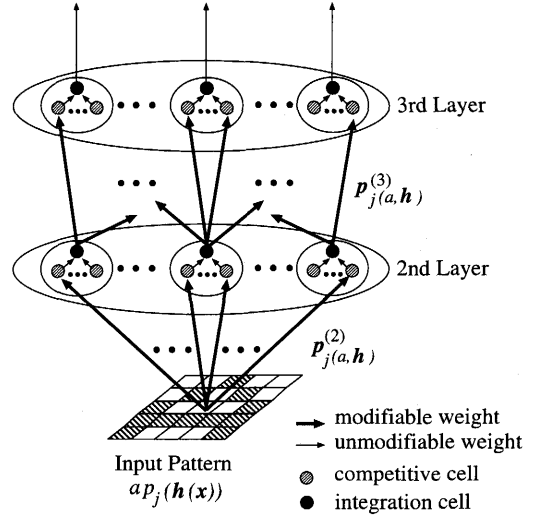


Fig. 2 Three layered competitive net, where $ap_j(\mathbf{h}(\mathbf{x}))$, $\mathbf{p}_j^{(2)}(a, \mathbf{h})$ and $\mathbf{p}_j^{(3)}(a, \mathbf{h})$ represent the signals appearing through the recognition process.

を出力し、テストパターン $ap_j(\mathbf{h}(\mathbf{x}))$ の入力時にはこのパターンをそのまま離散化した

$$\mathbf{p}_{j(a,h)}^{(2)} = a(p_j(\mathbf{h}(x_1)), p_j(\mathbf{h}(x_2)), \dots, p_j(\mathbf{h}(x_{N^2})))^T \quad (17)$$

を出力とする。

次に第 2 層の第 $i (\in I)$ クラスタは、そのクラスタ内の競合セルの荷重ベクトルの集合 $W_i^{(2)} = \{\mathbf{w}_{ij}^{(2)} \mid j \in K^{(2)}\}$ が第 i 番目の訓練パターン $p_i(\mathbf{g}(\mathbf{x}))$ に対する第 1 層の出力ベクトルを正規化した集合 $P_i^{(2)} = \{\mathbf{p}_{ig}^{(2)} / \|\mathbf{p}_{ig}^{(2)}\| \mid \mathbf{g} \in G\}$ にほぼ等しくなるように学習する (学習の詳細は 4 を参照のこと)。ここで $K^{(2)} = \{1, 2, \dots, |K^{(2)}|\}$ は第 2 層の各クラスタ内の競合セルの番号の集合である。すると学習後のネットにテストパターン $ap_j(\mathbf{h}(\mathbf{x}))$ が入力されると、各競合セルは荷重ベクトル $\mathbf{w}_{ij}^{(2)} \in W_i^{(2)}$ と入力ベクトル $\mathbf{p}_{j(a,h)}^{(2)}$ との内積を計算し、統合セルがそれらの内積の最大値を出力する。この計算は (5) 式を離散化したものと見ることができるので、結局、第 2 層は (9) 式にほとんど等しいベクトル

$$\mathbf{p}_{j(a,h)}^{(3)} \simeq \mathbf{s}_{j(a,h)} \quad (18)$$

を出力する。

第 3 層では、その第 $j (\in J)$ クラスタ内の競合セルの荷重ベクトル $\mathbf{w}_{jk}^{(3)}$ ($k \in K^{(3)}$) の 1 つが、変換していない第 j テストパターン $p_j(\mathbf{x})$ に対する第 3 層への入

カベクトル $\mathbf{p}_j^{(3)} \equiv \mathbf{p}_{j(l,e)}^{(3)}$ を正規化した $\mathbf{p}_j^{(3)}/\|\mathbf{p}_j^{(3)}\|$ を記憶する。ここで $K^{(3)}$ は第3層の各クラスタ内の競合セルの番号の集合である。すると(10)式より

$$\frac{\mathbf{p}_{j(a,h)}^{(3)}}{\|\mathbf{p}_{j(a,h)}^{(3)}\|} \simeq \frac{\mathbf{s}_{j(a,h)}}{\|\mathbf{s}_{j(a,h)}\|} \simeq \frac{\mathbf{s}_j}{\|\mathbf{s}_j\|} \simeq \frac{\mathbf{p}_j^{(3)}}{\|\mathbf{p}_j^{(3)}\|} \quad (19)$$

が成り立つので、学習後のネットに第 j テストパターンの変換パターン $ap_j(\mathbf{h}(\mathbf{x}))$ が入力されると、すべてのクラスタ内の正規化された荷重ベクトルのなかで第3層への入力パターン $\mathbf{p}_{j(a,h)}^{(3)}$ との内積が最大になるのは $\mathbf{p}_j^{(3)}/\|\mathbf{p}_j^{(3)}\|$ を記憶している第 j クラスタ内の荷重ベクトルになる。よって内積の最大値をもつクラスタの番号をネットが認識するパターン番号であると解釈すると上の過程により訓練した座標変換に対する不変認識ができることになる。また前述したように実際の文字認識などで異なる書体の文字 $ap_k(\mathbf{h}(\mathbf{x}))$ を第 j パターンとして認識させるには、第3層の第 j クラスタ内に $\mathbf{p}_{k(a,h)}^{(3)}/\|\mathbf{p}_{k(a,h)}^{(3)}\|$ を記憶させればよい。

4. 学習アルゴリズム

本研究ではネットの学習アルゴリズムとして LVC2 (Learning Vectors in Classes)¹⁰⁾ を用いる。以下にそのアルゴリズムを文献10)では明らかでなかった、エネルギー関数との関係から説明する。

ここでは第 L 層の第 i クラスタでの学習を考えると、層番号 L およびクラスタ番号 i を省略して説明する。ここで $L=2$ または $L=3$ であり $L=2$ のとき $i \in I$, $L=3$ のとき $i \in J$ である。さらに記憶すべきベクトルの集合を $P_c = \{\mathbf{p}_g \mid \|\mathbf{p}_g\| = 1, g \in G\}$ と表す。ここで G は記憶すべきベクトルの添字集合である。これは3の議論から座標変換以外のパターンも含む必要があるためここでは座標変換に限らないことにする。さて各クラスタ内の荷重ベクトル \mathbf{w}_k ($k=1, 2, \dots, |K|$) の個数 $|K|$ は記憶すべきベクトルの集合 P_c の要素数 $|P_c| = |G|$ より一般に小さく、すべての変換パターンを記憶することはできない。そこでエネルギー関数

$$E = \sum_{k \in K} \sum_{g \in G} \frac{1}{2} \|\mathbf{p}_g - \mathbf{w}_k\|^2 f(g) \quad (20)$$

を最小化することを考える。ここで G_k は集合 G を \mathbf{w}_k に関して次のように分割したもの、すなわち \mathbf{p}_g に最も近い荷重ベクトルが \mathbf{w}_k となる $g \in G$ の集合:

$$G_k = \left\{ g \mid \mathbf{w}_k = \underset{\mathbf{w}_m}{\operatorname{argmin}} \|\mathbf{p}_g - \mathbf{w}_m\|, g \in G, m \in K \right\} \quad (21)$$

である。さらに $f(g)$ は g が現れる確率である。ここで、もしセル数が十分大きく $|K| \geq |G|$ であり、すべ

ての $g \in G$ に対し $\mathbf{w}_m = \mathbf{p}_g$ となる \mathbf{w}_m ($m \in K$) が存在すればエネルギー E を零にできることを注意しておく。(20)式より荷重変化に対するエネルギーの勾配は

$$\frac{\partial E}{\partial \mathbf{w}_k} = -(\mathbf{p}_g - \mathbf{w}_k) f(g) \quad (22)$$

となる。従って勾配法または確率近似法 (例えば文献13) 参照) により、

$$\frac{d\mathbf{w}_k}{dt} = -\frac{\partial E}{\partial \mathbf{w}_k} \quad (23)$$

の解すなわち E の極小値に到達するための次の手順が与えられる。すなわち時刻 $t=1, 2, \dots$ において、学習のための入力 $\mathbf{p}_g(t)$ が与えられたとき、それに最も近い荷重 $\mathbf{w}_k(t)$ を選び

$$\mathbf{w}_k(t+1) = \mathbf{w}_k(t) + \beta_k(t)(\mathbf{p}_g(t) - \mathbf{w}_k(t)) \quad (24)$$

により変更すればよい。ここで $\beta_k(t)$ は小さな定数、あるいは t の増加とともに零に収束する変数とする。LVC2 では記憶すべき $\mathbf{p}_g(t)$ が正規化されていることから $\mathbf{w}_k(t)$ も正規化して

$$\mathbf{w}_k(t+1) = \frac{\alpha_k \mathbf{w}_k(t) + \mathbf{p}_g(t)}{\|\alpha_k \mathbf{w}_k(t) + \mathbf{p}_g(t)\|} \quad (25)$$

とする。ここで $\alpha_k(t) = av_k(t) + b$ であり、 a, b は定数、 $v_k(t)$ はこの荷重が時刻 t までに変更された回数であり、これをこの荷重の価値ともいう。以下の実験では、アフィン変換パターンの学習と認識¹⁰⁾ に有効であった $a=1, b=1$ を用いる。なお(25)式を α_k が十分大きいとして線形化すると(24)式を得る。

さて、以上の勾配法に基づいて得られるアルゴリズムでは局所解に陥る恐れがある。そこで、LVC2 では次の手法を用いる。すなわち $\mathbf{p}_g(t)$ に最も近い荷重を $\mathbf{w}_k(t)$ とし、さらに $\mathbf{w}_k(t)$ に最も近い荷重を $\mathbf{w}_m(t)$ とし、以下の処理を行う。

- (1) もし $v=0$ のセルがあれば入力ベクトルをその荷重ベクトルとし、 $v=1$ とする。 $v=0$ のセルがなくなれば(2)に移る。
- (2) もし $\|\mathbf{p}_g(t) - \mathbf{w}_k(t)\| \leq \|\mathbf{w}_m(t) - \mathbf{w}_k(t)\|$ ならば(25)式により $\mathbf{w}_k(t)$ を変更する。
- (3) そうでないならば、それまで最も利用されていない、すなわち最小の価値 $v_i(t)$ をもつ荷重を $\mathbf{w}_i(t+1) = \mathbf{p}_g(t)$ とし、またその価値は $v_i(t+1) = v_i(t)$ とする。

以上の処理の意味と役割は文献10) を参照のこと。

5. 計算機シミュレーション

5.1 使用するパターンセット

Microsoft Windows 98 に付属のフォント (Times New Roman) を用いて Fig. 3 に示すパターンセット

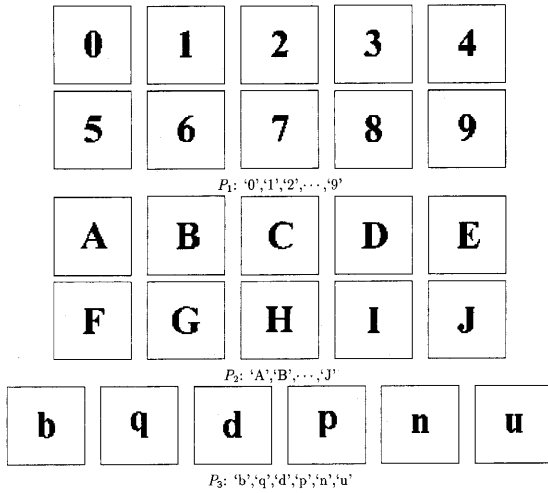


Fig. 3 Pattern sets.

を構成した。 P_1 は'0'から'9'の数字パターンセット、 P_2 は'A'から'J'のパターンセットである。 P_3 は回転すると等しい'b'と'q'、'd'と'p'および'n'と'u'のパターンセットである。これらのパターンは、各フォントの画像を 60×60 画素の背景の中央にはめ込み2値画像(0または1)としたものである。

5.2 射影変換に対する不変認識

5.2.1 第2層出力の不変性

まず射影変換パターンに対する第2層出力の不変性を得るため、射影変換行列

$$M = \begin{pmatrix} 1 + \Delta_a & \Delta_b & \Delta_e \\ \Delta_c & 1 + \Delta_d & \Delta_f \\ \Delta_p & \Delta_q & 1 \end{pmatrix} \quad (26)$$

とその逆変換行列を

$$\begin{aligned} |\Delta_a| \leq 0.1, & \quad |\Delta_b| \leq 0.1, & \quad |\Delta_e| \leq 2, \\ |\Delta_c| \leq 0.1, & \quad |\Delta_d| \leq 0.1, & \quad |\Delta_f| \leq 2, \\ |\Delta_p| \leq 0.03, & \quad |\Delta_q| \leq 0.03, \end{aligned} \quad (27)$$

の範囲でランダムに生成して P_1 の各パターンについてそれぞれ1000通りの変換パターンを作成し、第2層の対応するクラスタにそれぞれ学習させた。第2層の各クラスタの競合セルの数はそれぞれ200とした。

この後、テストパターンとして(26)式の変換行列 M の Δ_p と Δ_q の値をそれぞれ-0.03から0.03まで0.005間隔で変化させ、これら以外のパラメタの変化分は0に固定して P_2 の各パターンの射影変換パターンを生成した。 P_2 の'A'から生成されたテストパターンの一部をFig. 4(a)~(d)に示し、これらを含む変換パターンに対する第2層の出力結果を(e)に示す。この結果は第2層で学習された P_1 の変換パター

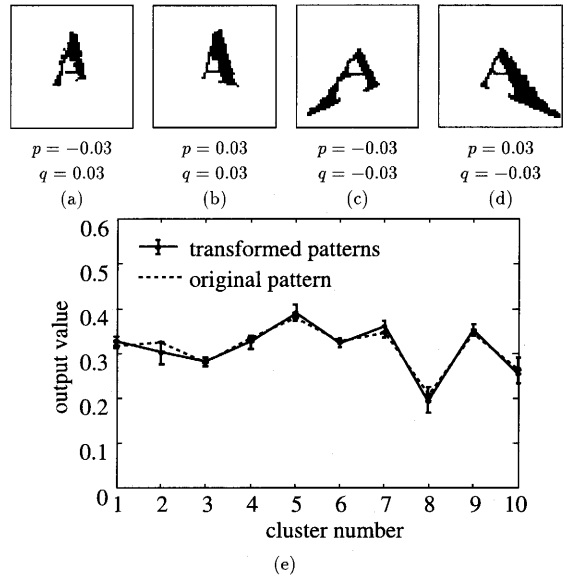


Fig. 4 (a)-(d) Transformed patterns and the parameters of projective transformations applied to 'A' in P_2 , and (e) invariance.

ンに施された変換を未学習のテストパターン'A'に施してネットに入力すると、第2層の出力はほぼ一定となるが、完全な不変性は得られていないことを示す。この不完全な不変性は、(8)式の a_h の*i*に対するばらつきや、実際のネットは連続なパターンや変換群を記憶できないために生ずる離散化誤差などに起因すると思われる。テストパターン'A'以外のパターンに対しても同様に不完全な不変性が得られるが、これらの不変認識への影響は次節で検討する。

5.2.2 不変認識

変換を施していない P_2 の各パターンを、前節5.2.1で訓練した第2層をもつネットの第3層の対応するクラスタにそれぞれただ一度ずつ学習させた。次に、 P_2 の各パターンに対して前節のテストパターンの生成方法と同様に Δ_p, Δ_q を変化させて生成した射影変換パターンを用いて認識テストを行った。全テストパターンに対する認識率は95.0%となり、より詳細な認識率としてFig. 5の(a)と(b)が得られた。(a)はパターンによって認識率が異なることを示し、(b)は Δ_p と Δ_q の絶対値が大きいとき認識率が低いことを示す。これら結果はそれぞれ次の2つの誤認識の原因と対策を示唆する。1つの原因は訓練パターンとテストパターンの関係に起因している。すなわち2.1の議論によると、各テストパターンに非常に類似した訓練パターンを記憶すると、そのテストパターン

に対する識別能力が得られるので、逆に任意に選んだ訓練パターンを用いると各テストパターンの認識率が Fig. 5(a) のようにそれぞれ異なることが理解できる。このことを考慮し、第2層への訓練パターンセットとして P_1 の代わりにテストパターンセット P_2 を用いて上と同様の実験を行った結果、すべてのパターンが認識できた。この結果はテストパターンと訓練パターンを等しくすることにより得られたものであり、不変認識におけるネットの高い学習能力を示している。

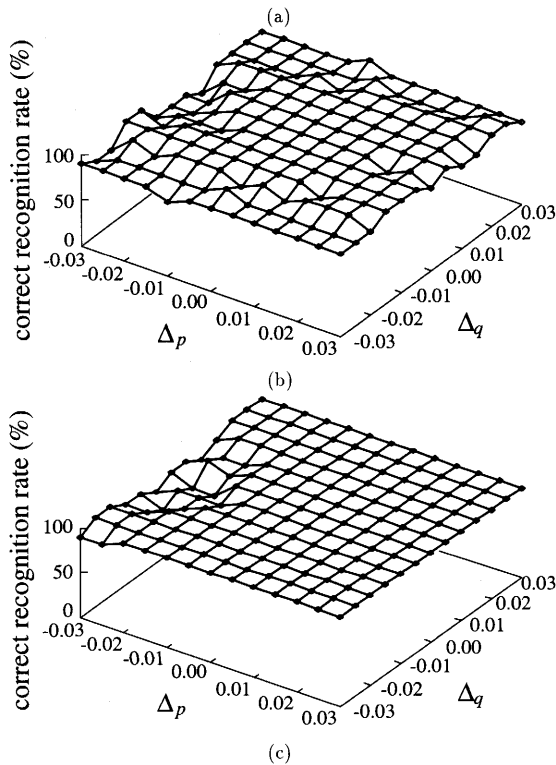
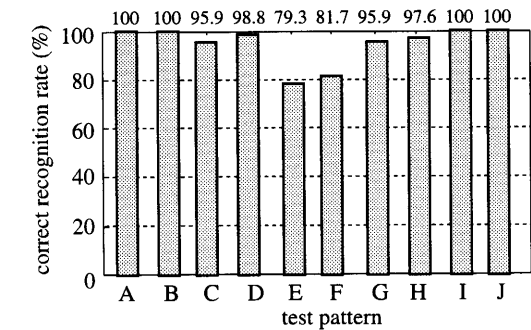


Fig. 5 Correct recognition rates for (a) the transformed patterns, (b) the parameters Δ_p and Δ_q , and (c) those after the training for $\Delta_p = \pm 0.02$ and $\Delta_q = \pm 0.02$.

もう1つの原因は前節で示した第2層出力の不完全な不変性である。特に射影変換行列の Δ_p と Δ_q の絶対値が大きくなると(8)式の近似誤差が大きくなるので Fig. 5(b) の結果が理解できる。これらの誤認識は、2.1の議論から、大きな Δ_p と Δ_q に対する第2層の出力も第3層で学習することにより除去できると考えられる。そこで第3層に $(\Delta_p, \Delta_q) = (\pm 0.02, \pm 0.02)$ を用いた4通りの組合せに対する P_2 の変換パターンを追加学習させ認識テストを行った結果、Fig. 5(c) が得られ全体の認識率は 98.5% に増加した。さらにそこで誤認識が生じた $(\Delta_p, \Delta_q) = (-0.025, 0)$ および $(-0.03, -0.03)$ に対する変換パターンも追加学習すると、すべてのパターンが正しく認識できた。結局、第3層で6通りの Δ_p と Δ_q の組合せに対するテストパターンだけの追加学習を行うことですべてのテストパターンを認識できたことになり、これはネットの高い汎化能力を示すといえる。

5.3 回転変換に対する TC 汎化能力

提案するネットは TC 汎化能力、すなわち訓練パターンに内在する変換群のみに対する不変性を獲得する能力をもつ。ここでは回転を含まない変換群と回転を含む変換群を用いた P_1 の変換パターンを学習させ、回転して位置を調整すると等しくなるパターンを含む P_2 を用いてテストすることによりこの能力を検討する。

まず回転を含まない変換群として(26)式の変換行列 M の各パラメータを

$$\begin{aligned}
 &|\Delta_a| \leq 0.1, \quad |\Delta_b| \leq 0.1, \quad |\Delta_e| \leq 2, \\
 &|\Delta_c| \leq 0.1, \quad |\Delta_d| \leq 0.1, \quad |\Delta_f| \leq 2, \\
 &|\Delta_p| \leq 0.01, \quad |\Delta_q| \leq 0.01,
 \end{aligned}
 \tag{28}$$

の範囲でランダムに生成した。実際にはこの範囲の変換行列 M により構成される射影変換群は -5.74° から $+5.74^\circ$ までの回転変換を含むが、この範囲は離散化や入力パターン間の回転中心のずれなどの影響を除くためのものであり、この節では上式の範囲により生成される変換群を非回転変換群と呼ぶことにする。これに対し回転を含む変換群は上式の範囲の M と回転角 θ が $0^\circ \leq \theta < 360^\circ$ の範囲のランダムな値をとる回転変換 M_θ を合成した

$$MM_\theta = M \begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix}
 \tag{29}$$

により生成した。

第2層には、上記の M と MM_θ をそれぞれ用いて 5.2.1 と同様の方法でパターンセット P_1 の各パターンを学習させた。以下では学習後のこれらのネットを

Table 1 Results of the recognition tests for rotations.

	非回転パターン 認識テスト	回転パターン 認識テスト
非回転パターン学習ネット	100.0%	47.6%
回転パターン学習ネット	50.0%	100.0%

それぞれ非回転パターン学習ネットおよび回転パターン学習ネットと呼ぶ。次にこれらのネットの第3層に P_3 の各パターンをそれぞれ別のクラスタに1回ずつ学習させた場合と、 P_3 の'b', 'd', 'n'だけをそれぞれ別のクラスタに1回ずつ学習させた場合について認識テストを行った。前者は P_3 の各パターンの変換パターンをネットに入力したとき第3層の対応するクラスタのセルが最大値を出力した場合を正しい認識とし、後者は'b'と'q', 'd'と'p', 'n'と'u'の変換パターンをネットに入力したときそれぞれ'b', 'd', 'n'に対応する第3層のクラスタのセルが最大値を出力した場合を正しい認識とする。以下ではこれらをそれぞれ非回転パターン認識テストおよび回転パターン認識テストと呼ぶ。なお非回転パターン認識テストは離散化などに起因するわずかな回転変動に対してもうまくいくことを確かめるため、 P_3 の各パターンについて -5° から 5° までの範囲で 1° 間隔で生成した回転を含む変換パターンを用いて行った。また回転パターン認識テストは -180° から 180° までの範囲で 1° 間隔で生成した回転を含む変換パターンを用いて行った。Table 1は以上の認識テストの結果を示す。この表により、学習した変換群のみに対して高い不変認識能力、すなわちTC汎化能力が得られていることが分かる。

6. む す び

本論文では線形および非線形座標変換に対する不変認識能力を獲得するための手法を示し、それを実現する多層競合ネットを構成した。さらに非線形座標変換の具体例として射影変換を挙げ、これに対する不変認識能力およびTC汎化能力を示した。

本論文で構成したネットは異なる書体の文字が混合される場合の文字認識や手書き文字認識¹⁾などへも応用できることが2.1の議論から理解できる。このとき、より高い認識能力や汎化能力を得るには使用する訓練パターンの数と種類、訓練する変換の種類などをうまく選ぶ必要があるが、これらについての考察や実験は今後の課題である。さらに、射影変換以外の非線形座標変換に対する不変認識、例えば、ビンのラベル、ボールの表面に描かれた文字パターン、衣服などの柔らかく変形しやすい物体の表面の文字パターンな

どの不変認識への応用も考えられる。このとき記憶すべき訓練パターンを生成するための Jacobian $|g_x|$ は考慮する座標領域 D 内の主な点における変換前後の微小面積の比を(4)式により計算すればよいと考えられるが、この手法の検討と実際の場面への応用実験も今後の課題である。

参 考 文 献

- 1) Fukushima, K. (1982): Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position, Pattern Recognition, Vol. 15, No. 6, pp. 455-469
- 2) Widrow, B., Winter, R. G. and Baxter, R. A. (1988): Layered neural nets for pattern recognition, IEEE Trans. Acoust., Speech Signal Process., Vol. 36, No. 7, pp. 1109-1118
- 3) Giles, C. L. and Maxwell, T. (1987): Learning, invariance, and generalization in high-order neural networks, Applied Optics, Vol. 26, No. 23, pp. 4972-4978
- 4) Perantonis, S. J. and Lisboa, P. J. G. (1992): Translation, rotation, and scale invariant pattern recognition by high-order neural networks and moment classifiers, IEEE Trans. Neural Networks, Vol. 3, No. 2, pp. 241-251
- 5) Konen, W. K., Maurer, T. and Malsburg, C. (1994): A fast dynamic link matching algorithm for invariant pattern recognition, Neural Networks, Vol. 7, No. 6-7, pp. 1019-1030
- 6) Marr, D. (1982): Vision—A Computational Investigation into the Human Representation and Processing of Visual Information. New York: W. H. Freeman and Company
- 7) Hotta, K., Iwata, A., Matsuo, H. and Suzumura, N. (1992): A large scale neural network CombNET-II, IEICE Trans., Vol. J 75-D-II, No. 3, pp. 545-553
- 8) Kurogi, S. (1993): Competitive neural network for affine invariant pattern recognition, Proc. of IJCNN '93, pp. 181-184
- 9) Kanatani, K. (1990): Group-Theoretical Methods in Image Understanding. Berlin, Tokyo: Springer-Verlag
- 10) Ren, S. and Kurogi, S. (1998): Learning algorithms using firing numbers of weight vectors for WTA networks in rotation invariant pat-

- tern classification, IEICE Trans. FECCS, Vol. e 81-A, No. 1, pp. 175-182
- 11) Kurogi, S., Nishida, T. and Ren, S. (1998) : Recognition of hand-written characters by a multi-layered competitive net, Proc. of ICONIP '98, Vol. 2, pp. 1086-1089
 - 12) 山口富士夫 (1990) : コンピュータディスプレイによる図形処理工学. 東京 : 日刊工業新聞社
 - 13) Diamantaras, K. I. and Kung, S. Y. (1996) : Principal Component Neural Networks. New York : John Wiley & Sons, Inc.