

研究

並列分散処理による能動的画像認識モデル：顔の認識

津村 徳道・伊東 一良・一岡 芳樹

大阪大学工学部応用物理学教室 〒565 吹田市山田丘 2-1

(1994年6月27日受付, 1994年9月16日受理)

Active Vision by Parallel Distributed Processing: Face Recognition

Norimichi TSUMURA, Kazuyoshi ITOH and Yoshiki ICHIOKA

Department of Applied Physics, Faculty of Engineering,
Osaka University, 2-1, Yamadaoka, Suita 565

(Received June 27, 1994; Accepted September 16, 1994)

In this paper, we propose a new parallel distributed processing model for an active vision system. This model is assumed to have active functions of eye-movement and zooming. The present model has a relational module and four functional modules, the pattern-matching, face-identification, pattern-discrimination and camera-control modules. The relational module interconnects the functional modules, and infers the eye-movement in cooperation with the functional modules. The present model is constructed for the application to face recognition. In the experiment, the faces of 61 persons are recognized by the selective eye-movements of the proposed model.

1. はじめに

ものを見るとき、人間の目は、1秒間に約3回もの頻繁な視点の移動を行う¹⁾。人間は、得られた空間位置の情報²⁾や経験による知識等に基づいた視点の移動により、選択的に外界から視覚情報を獲得していると考えられる。知能ロボットにおいても、目的に応じた視覚情報を視覚センサーを動かして選択的に獲得することが必要である。こうした視点の移動による能動的な視覚システムは、1980年代後半から研究されている^{3,4)}。

一方、ニューラルネットワークとも呼ばれる並列分散処理モデル^{5,6)}がモデル化の困難な実世界で用いられる画像認識技術として注目され、いくつかのモデルが提案されている⁷⁻⁹⁾。しかし、現在のところ中野らのモデル¹⁰⁾を除き、視点の移動などを用いた能動的な画像認識モデルは、並列分散処理モデルの枠組みの中では実現されていない。また、中野らのモデルは視点の移動を行う点では能動的ではあるが、目的に応じた選択的な視点の移動は行っていない。

本論文では、並列分散処理により、選択的な視点の移動を行う能動的画像認識モデルを提案する。提案するモ

デルは、視野内のパターンをモデルの記憶するパターンと照合するモジュール、視覚センサーの制御を行うモジュール等の各種機能モジュールと、各モジュールを接続し、視点の移動による認識過程を推論する関係モジュールで構成されている。関係モジュールは、調和回路網¹¹⁾と同じ構造をしているが、調和回路網とは異なり、層間の方向性のある接続により順序の関係する推論を実現している。次の章で、提案するモデルについて述べ、3章で、計算機実験により、顔画像の認識実験を本モデルを用いて行った結果について述べる。

2. 並列分散処理による能動的画像認識モデル

本章では、顔の認識を例にとり、提案する並列分散処理による能動的画像認識モデルについて説明する。顔の認識だけを扱う場合、目、鼻等のように要素となるパターンの位置関係が、すべての認識対象に対してほぼ同一であると考えられる。

2.1 視点の移動とズーム操作による能動的画像認識モデル

Fig. 1(a)に、対象となる人の顔に関する480×480画素の通常の画像を示す。同じ人の顔から、視覚センサー

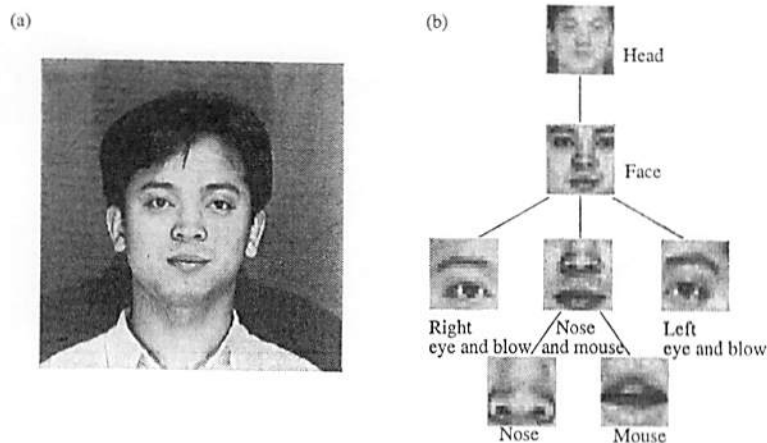


Fig. 1 Example of image of human face; (a) a usual image, (b) a pattern tree constructed from local patterns.

の移動とズーム操作を想定して切り出した特徴的な複数の局所パターン (local pattern) を Fig. 1(b)に示す。Fig. 1(b)に示す複数の局所パターンは、全体をまとめてパターン木 (pattern tree) と呼ぶ⁴⁾。この時、パターン木には視点の移動のための位置情報とズーム操作のための倍率の情報が付随しているものとする。パターン木による画像の表現方法は、通常、画像の特徴抽出等により獲得されるが⁴⁾、ここでは、認識処理を中心に論議するため、これらが与えられているものとして話を進める。Fig. 1(b)において、パターン木の上から1段目の局所パターンは、外界の視覚情報から切り出された頭部の画像である。頭部からさらに細かい構造を切り出したのが上から2段目の顔面の画像である。同様に、顔面からは右目・眉、鼻・口、左目・眉の三つの画像が、鼻・口からは鼻、口の二つの画像が切り出される。各局所パターンの画素数は、想定した視覚センサーの画素数で決まり、Fig. 1(b)では 17×17 画素である。想定した視覚センサーの画素数は非常に少ないが、小さな領域では高い解像度が得られる。物体をパターン木として記憶すると、画像認識システムの記憶容量を大幅に節約することが可能である。

提案するモデルでは、パターン木に関する知識を並列分散処理モデルを構成する処理ユニット (ニューロン) 間の信号結合強度の分布として記憶し、ユニット間の相互作用による並列分散処理により顔の認識を行う。

Fig. 2 に、今回実現した顔の認識モデルの基本構造を示す。四つの機能モジュール (functional modules) と、それらと相互に作用する関係モジュール (relational module) で構成されている。以下に、各モジュールの構造、役割について述べる。

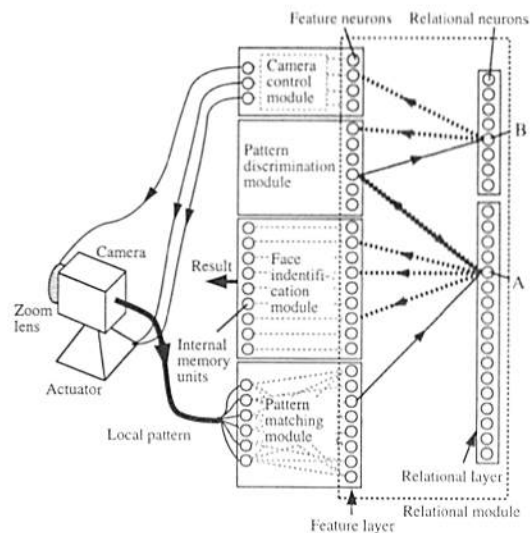


Fig. 2 Parallel distributed processing model for active vision.

パターン照合モジュール (pattern-matching module)

入力層、出力層からなる二層のニューラルネットワークである。出力層は関係モジュールの特徴層の一部を成す。本モジュールの入力層、出力層と関係モジュールを介したパターン識別モジュール内の一層で三層の RCE ネットワーク^{12,13)}を構成する。RCE ネットワークとは、学習により、その第二層においてニューロンの生成とニューロンのしきい値の修正が行われるニューラルネットワークである。学習後の RCE ネットワークは、第二層の各ニューロンで入力パターンとそのニューロンの持つ参照パターンとの照合を行う。その照合結果が第三層に伝えられ、入力パターンに対する識別結果が得られる。したがって、パターン照合モジュールは、局所パターン

を入力とし、モジュールの持つ複数の参照パターンとの照合結果を関係モジュールに伝える。また、ニューロンのしきい値は、学習に用いた一つの局所パターンに対して、出力層の一つの特徴ニューロンのみが活性化されるように修正されている。

顔同定モジュール (face-identification module)

認識すべき個人の顔一人一人に対応した特徴ニューロンと内部メモリユニットにより構成されている。この特徴ニューロンは、関係モジュールから局所パターンの照合結果を受け取る。特徴ニューロンの値と内部メモリユニットの値に基づき顔の同定を行う。認識すべき顔に対して、複数の候補があるために顔の同定ができない場合には、候補となる顔の中からどれかの顔を仮説として取り上げ、その結果を関係モジュールに伝える。各内部メモリユニットの値には、処理を始めてから現在までに、現在候補に上がっている顔が仮説として取り上げられた回数が記録されている。詳しい動作については、2.2で述べる。

パターン識別モジュール (pattern-discrimination module)

関係モジュールの特徴層の一部を成す一層のみで構成されている。関係モジュールから局所パターンの照合結果を受け取り、局所パターンが頭部、顔面、右目・眉等のクラスのうち、どのクラスに属するかを識別する。または、関係モジュールから次に視点を移すべき局所パターン (注意視野パターン) に関する情報を受け取り、そのクラスを識別する。各クラスに対し、一つの特徴ニューロンが対応するため、今回は七つの特徴ニューロンのみで構成される。

カメラ制御モジュール (camera-control module)

関係モジュールの特徴層の一部を入力層とし、カメラの位置、ズームレンズの調整のための信号を出力するモジュールである。入力層の特徴ニューロンの一つ一つは、特定の視点の移動とズーム操作を指示するユニットである。今回の実験では、実際には視点の移動やズーム操作は行わない。Fig. 1(b)において、局所パターンの間を結ぶ直線は6本ある。そのため、パターン木中での視点の移動は、全部で12種類必要である。したがって、カメラ制御モジュールは、12種の視点移動のための12個の特徴ニューロンを備えている。これらのうち一つが活性化した場合、それに固有の位置で得られる局所パターンをパターン照合モジュールに与える。

関係モジュール (relational module)

特徴ニューロン (feature neuron) の集まりである特

徴層 (feature layer) と、関係ニューロン (relational neuron) の集まりである関係層 (relational layer) の二層で構成されている。関係モジュールを構成するニューロンは、応答値 (response value) r 、推論値 (inferring value) i という二つの独立した内部値を持つ。各値は、顔同定モジュール中の特徴ニューロンを除き、すべて0または1の値を持つ。顔同定モジュール中の特徴ニューロンの推論値は、負の値を持ち得ることができる。ニューロンの応答値または推論値が1の場合、それぞれ、ニューロンが応答値、推論値に関して活性化しているという。認識処理において、応答値は実際に存在する入力からの信号を保持し、推論値は推論時に仮説により形成される信号を保持する。応答値は、各層のニューロン間を、定められた方向に伝搬する。これに対し推論値はその逆方向に伝搬する。

関係ニューロンには二つの種類が存在する。一つは、入力された局所パターンが、どの顔のどのクラスに属するかを記憶する関係ニューロン(A)である。Fig. 2には、パターン照合モジュールからこの関係ニューロンへ、この関係ニューロンから顔識別モジュールへの結合と、パターン識別モジュールとこの関係ニューロン間の双方向の結合が形成されている様子が示されている。もう一つは、視線の移動前の局所パターンのクラスから移動後のクラスへのカメラ制御に関する情報を記憶する関係ニューロン(B)である。図には視線の移動前のクラスを示すパターン識別モジュール中の特徴ニューロンからこの関係ニューロンへ、この関係ニューロンから移動後のクラスを示す特徴ニューロンとカメラ制御モジュールへの結合が形成されている様子が示されている。これらの関係ニューロンは、それぞれ独立に信号を受・発信する二つの野を構成している。前者の関係ニューロンで構成される野は、注意視野推論野 (field for inferring visual field of attention)、後者の関係ニューロンで構成される野は視点制御推論野 (field for inferring eye-movement) と呼ぶことにする。

2.2 認識動作におけるモジュール間の協調

本節では、モデル全体の認識動作を説明する。関係モジュールと、内部メモリユニットのみを示した Fig. 3を用いて説明する。各層の状態は、応答値ベクトル、推論値ベクトルで表される。以下に、説明で用いられるパラメータや関数等の定義を示す。

パラメータ、関数等の定義

$r^{(F)}, i^{(F)}$: 特徴層の応答値、推論値ベクトル。

$r^{(RA)}, i^{(RA)}$: 注意視野推論野の応答値、推論値ベ

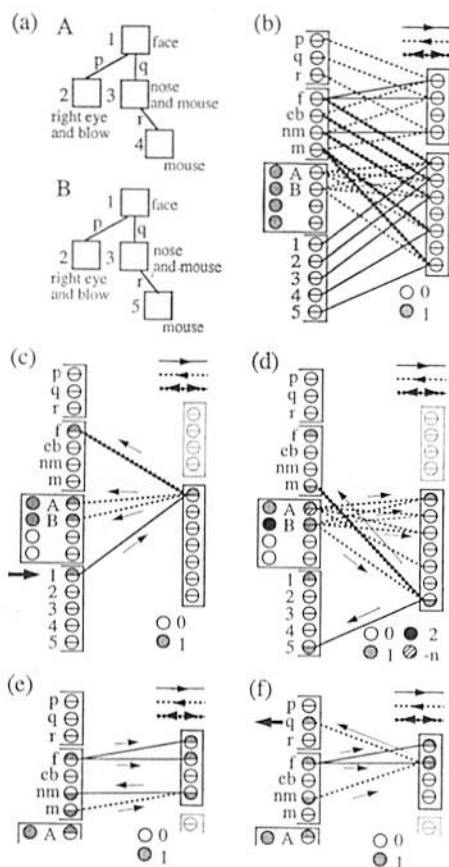


Fig. 4 Example of processes of face recognition; (a) pattern trees, (b) recorded interconnections, (c) face-identification, (d) detection of visual field of attention, (e) process of deciding eye-movement, (f) decided eye-movement.

程(1)に従い、顔同定モジュールのA, Bの特徴ニューロンが応答値に関して活性化される。顔同定モジュールにおいて複数の特徴ニューロンが活性化している($\|r^{(F)}\| > 1$) ため次の認識過程に進む。

認識過程(2) 最大の値をもつ内部メモリユニットの中の一つの番号を j とする。この時、対応する番号 j の顔同定特徴ニューロンの推論値 $i_j^{(F)}$ に1を与えることにより、対象とする顔が j 番目の特徴ニューロンが示す顔であるという仮説を設ける。同時に、内部メモリユニット j の値 m_j を1だけ大きくする。仮説からはずれた応答値が正の特徴ニューロンの推論値には、負の値 $-n$ を与える。この時、 n は顔同定モジュール内の特徴ニューロン数より大きい値とする。応答値が0の特徴ニューロンの推論値には、0を与える。これらを含む推論値ベクトル $i^{(F)}$ が、次式に従い関係層に伝搬し、 $i^{(RA)}$ が得られる。

$$i^{(RA)} = \text{Step}(W^{(RA-F)} i^{(F)}). \quad (5)$$

注意視野推論野の推論値ベクトルのノルム $\|i^{(RA)}\|$ の値に従い、以下の視点を移すべき注意視野パターンの検出処理が行われる。

注意視野パターンの検出処理

(a) $\|i^{(RA)}\| > 0$ の場合、注意パターンが検出され、 $i^{(RA)}$ が次式に従い、パターン識別、パターン照合モジュールに伝搬し、次式のように $i^{(F)}$ が得られる。

$$i^{(F)} = \text{Step}(W^{(F-RA)} i^{(RA)}). \quad (6)$$

$i^{(F)}$ が得られた後、この認識過程を終了し、次の認識過程(3)に進む。

(b) $\|i^{(RA)}\| = 0$ の場合、注意視野パターンが検出されるように仮説の拡張が行われる。現在、推論値が $-n$ で、内部メモリユニットの値が最も大きい特徴ニューロン j の推論値 $i_j^{(F)}$ に1を与え、再び、式(5)に基づく信号の伝搬、注意パターンの検出処理を行う。この時、 m_j の値を1だけ大きくする。

Fig. 4(d)では、本認識過程に従い、ニューロンAの推論値に $-n$ 、ニューロンBの推論値に1を与えている様子が示されている。また、注意視野パターンの検出処理により、次に視野に捉えるべきクラスの特徴ニューロン m と、その局所パターンに対応する特徴ニューロン5の推論値が活性化されている。この例では、対象とする顔の候補が二つしかないため、一度の仮説により、注意パターンが検出されている。しかし、実際には3章の実験で示すように、一つの局所パターンに対して顔の候補は複数あり、仮説の拡張は頻繁に行われる。

認識過程(3) 特徴層の応答値、推論値ベクトル $r^{(F)}$, $i^{(F)}$ が、それぞれ次の(7), (8)式に従い伝搬し、 $r^{(RE)}$, $i^{(RE)}$ が得られる。

$$r^{(RE)} = \text{Step}(W^{(F-RE)} r^{(F)}), \quad (7)$$

$$i^{(RE)} = \text{Step}(W^{(F-RE)} i^{(F)}). \quad (8)$$

この応答値ベクトル $r^{(RE)}$ は、 $(r^{(RE)})'$ を修正後のベクトルとしたとき、

$$(r^{(RE)})' = r^{(RE)} \otimes i^{(RE)}, \quad (9)$$

に修正される。修正後の応答値ベクトルのノルム $\|(r^{(RE)})'\|$ の値に従い、以下のように注意視野パターンを捉えるために最初に行う視点の動作に関する決定処理が行われる。

視点動作の決定処理

(a) $\|(r^{(RE)})'\| > 0$ の場合、応答値ベクトル $(r^{(RE)})'$ が、次の(9)式に従い特徴層のカメラ制御モジュールに伝搬し、 $r^{(F)}$ が得られる。

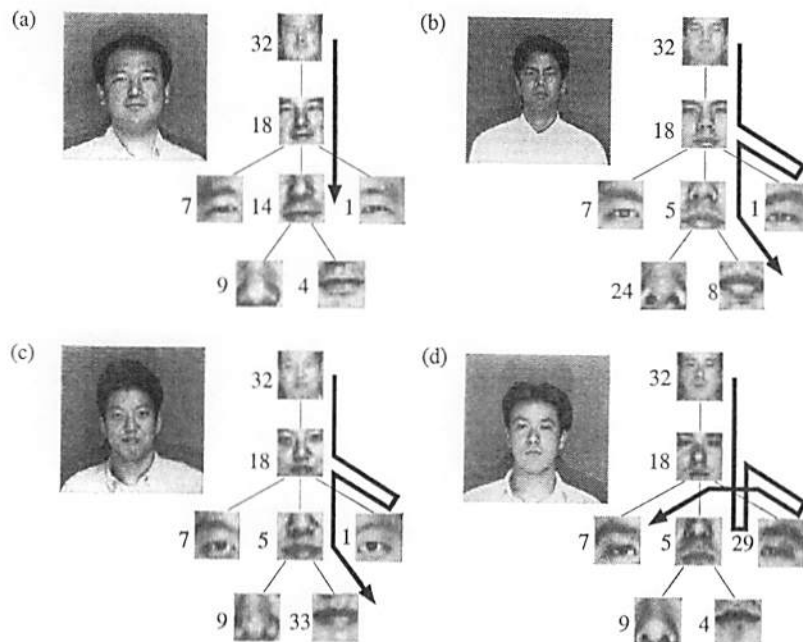


Fig. 5 Selective eye-movements for face recognition.

$$r^{(F)} = \text{Step}(W^{(RE-F)}(r^{(RE)})^t). \quad (10)$$

$r^{(F)}$ に従いカメラの制御が行われ、次の局所パターンを捉える。認識過程(1)からの認識処理を続ける。

(b) $\|r^{(RE)}\| = 0$ の場合、視点動作が決定されず、推論値ベクトル $i^{(RE)}$ が、次式に従い特徴層に伝搬し、 $i^{(F)}$ が得られる。

$$i^{(F)} = \text{Step}(W^{(F-RE)}i^{(RE)}). \quad (11)$$

これをもとに、再びこの認識過程(3)を繰り返す。

Fig. 4(e)の状態では、視点動作の決定処理の際、視点制御推論野において応答値、推論値両方に関して活性化している関係ニューロンが存在しない。したがって、(9)式の左辺の $(r^{(RE)})^t$ のノルム $\|r^{(RE)}\|$ は 0 となり、上記(b)のとおり、(11)式による推論値ベクトルの伝搬が行われている。Fig. 4(f)では、2度目の視点動作の決定処理の結果、応答値、推論値両方に関して活性化している関係ニューロンが存在する ($\|r^{(RE)}\| > 0$)。そのため、(10)式による応答値ベクトルの伝搬が行われ、カメラ制御モジュールの特徴ニューロン q が発火し、選択された注視視野パターンへ向かう視点の移動が行われる。

この例ではその後、視点の移動により得られた局所パターンをもとに、上記と同じ認識過程(1)、(2)、(3)を経て、口の部分への視点の移動を行う。その結果、顔同定モジュールにおいてAに関するユニットの活性化の

みが残し、対象はAと認識される。

3. 計算機実験

3.1 パターン木の収集とパターン照合モジュールの学習

男子 60 名、女子 1 名、計 61 名の平常顔をビデオカメラにて撮影した。Fig. 1(a)のような顔画像を 61 枚用意する。単純なパターンマッチング技術により、一人につき 7 個の白黒の局所パターンからなる Fig. 1(b)のようなパターン木を得る。この時、局所パターンでは、画素値の平均、分散に関する規格化が行われる。こうして得られたパターン木をもとに、提案するモデルの有効性を確認する。

RCE ネットワークの学習により形成されたパターン照合モジュールの特徴ニューロン数は、49 個であった。RCE ネットワークは、右目・眉、鼻等の異なるクラスに属する局所パターンに対しては、必ず異なる特徴ニューロンが形成されるように学習した。また、同じクラスであるが異なる顔の局所パターンの間については、特徴ニューロンのしきい値に下限を設定し、特徴ニューロンの共有を可能にした。しきい値の下限値は、異なる顔の間で、七つの局所パターンに対する特徴ニューロンをすべて共有しないという条件のもとで、全体として特徴ニューロン数が最小になるように人為的に設定した。生成された特徴ニューロンの内訳は、頭部の局所パターンに

対しては14個、顔面6個、左目・眉8個、右目・眉5個、鼻・口3個、鼻9個、口4個である。

3.2 実験結果

本モデルでは、視点制御推論野の12個の関係ニューロンにより12種の視点移動を、注視パターン推論野の49個の関係ニューロンにより61人の顔の特徴を記憶している。本モデルにより、記憶した顔画像を選択的な視点の移動により認識した。Fig. 5(a)~(d)に、4名の顔に関して、通常の画像、パターン木、認識が完了するまでの視点移動の軌跡を示す。各局所パターンの左側に、パターン照合モジュールにおいてその入力に対して活性化される特徴ニューロンの番号を記した。Fig. 5(a)の顔は、2度目の視点移動で選択的に鼻・口のパターンに視点を移動している。これは、注意視野パターンの検出により、Fig. 5(a)とFig. 5(b)~(d)とを区別するクラスとして鼻・口が検出されたためである。Fig. 5(b)と(c)の顔は、通常の画像を見てもわかるとおり、口元を除き非常によく似ている。その結果、最終的に口のパターンまで視点の移動を行った。一度、右目・眉に視点を移動しているのは、Fig. 5(d)の顔でないことを確認するためである。Fig. 5(d)は、無駄な視点の移動をしているように見えるが、これはFig. 5に表さなかった複数の顔が、Fig. 5(d)と大部分の局所パターンで似ているために行われたものである。

本モデルにより、61名に関して顔を認識するまでの視点移動の回数は平均2.89回であった。それに対して、視点の移動に基づき、局所パターンの全数検査(縦型探索¹⁴⁾)により顔の認識を行った場合の視点移動の回数は平均4.05回であった。並列分散処理による推論により、効率の良い視点の移動が行われている。

4. おわりに

本研究では、カメラの移動とズーム操作による能動的画像認識のための並列分散処理モデルを提案した。提案するモデルを実際の顔の認識に適用し、知識をもとにした選択的な視点の移動が行われることを確認した。

一般に用いられる並列分散処理モデルでは、すべての入力情報はあらかじめ与えられている。本モデルでは、パターン木という階層的な知識源から視点の移動により逐次的に情報を集める。そのため、三つの認識過程に基づく逐次的な認識を行う。階層的な知識を、単純な二層構造で記憶できるため、どのレベルにおいても全く同じ三つの認識過程が用いられる。階層の順序に関する知識は、方向性のある結合で保持されている。全体と部分の

関係で記述できる対象に対して、このような階層間での時間分割処理が有効であることは、Hintonによっても主張されている¹⁵⁾。

今回のモデルの構築は、学習は行わず、簡単なパターンマッチング技術によって既に得られた顔のパターン木を記憶することにより行った。したがって、視点の移動は、その記憶と現在の入力パターンが何(what)であるかの情報のみで行われる。局所パターンの位置関係の多様な物体の認識に本モデルを適用するとき、記憶された視点の移動先には対象とする局所パターンが存在しない場合が生ずる。生理学的にも確認されているように²⁾、入力パターンから視点を移すべき局所パターンがどこ(where)に存在するかの情報を得る機能モジュールを本モデルに加えることが、本モデルの多様な物体認識への適用や学習の導入のためには必要である。

本モデルでは、選択的な視点の移動のみに焦点を当て、表情の変化した顔等の、記憶した顔画像以外の入力に対する汎化能力あるいはロバスト性を考慮しなかった。しかし、パターン照合モジュールは、各局所パターンに対して汎化能力を持つと考えられる。今後、パターン照合モジュールの能力向上とともに、一部の局所パターンが得られなかった時等の画像全体に対する汎化能力についても考慮する必要がある。

文 献

- 1) 池田光男: 目は何を見ているのか—視覚系の情報処理—(平凡社, 1988).
- 2) M. Mishkin, L.G. Ungerleider and K.A. Macko: "Object vision and spatial vision: two cortical pathways," *Trends In Neurosciences*, **6** (1983) 414-417.
- 3) R. Bajcsy: "Active perception," *Proc. IEEE*, **76** (1988) 996-1005.
- 4) P.J. Burt: "Smart sensing within a pyramid vision machine," *Proc. IEEE*, **76** (1988) 1006-1015.
- 5) D.E. Rumelhart, J.L. McClelland and PDP Research Group: *Parallel Distributed Processing* (MIT Press, Cambridge, 1986).
- 6) G.E. Hinton, ed.: *Connectionist Symbol Processing* (MIT Press, Cambridge, 1991).
- 7) K. Fukushima: "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biol. Cybernetics*, **36** (1980) 193-202.
- 8) Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard and L.D. Jackel: "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, **1** (1989) 541-551.
- 9) J.E. Hummel and I. Biederman: "Dynamic binding in neural network for shape recognition," *Psychol. Rev.*, **99** (1992) 480-517.
- 10) K. Nakano, M. Niizuma and T. Omori: "Model of neural visual system with self-organization cells,"

- Biol. Cybernetics, **60** (1989) 195-202.
- 11) P. Smolensky: "Information processing in dynamical systems: Foundations of harmony theory," *Parallel Distributed Processing, vol. 1*, ed. D.E. Rumelhart, J.L. McClelland and PDP Research Group (MIT Press, Cambridge, MA, 1986) pp. 194-281.
 - 12) D.L. Reilly, L.N. Cooper and C. Elbaum: "A neural model for category learning," Biol. Cybernetics, **45** (1982) 35-41.
 - 13) 黒川忠由: "ニューラル・ネットワークとその開発支援ソフト", インターフェース, **149** (1989) 239-250.
 - 14) 長尾 真: 知識と推論 (岩波書店, 東京, 1988).
 - 15) G.E. Hinton: "Mapping part-whole hierarchies into connectionist networks," *Connectionist Symbol Processing*, ed. G.E. Hinton (MIT Press, Cambridge, 1991) pp. 47-75.