知覚と認知の計算理論

乾　敏郎

人間を含め高等動物の持つ高度な視覚機能がいかにして神経回路網で実現されているかを明らかにすることが視覚研究の究極の目標である。「なぜ見えるように見えるのか」という疑問は誰もが持つ疑問であるが、これでは問題がうまく定式化されていない。視覚研究は長年にわたり実験科学として多くの知識を集積してきた。しかし視覚の目的を明確に定式化し、視覚計算論という新しい分野を切り開いた David Marr(1982) の出現はこの一〇年間に視覚研究を大きく変貌させたのである。また八〇年代は高次脳機能に関する解剖学や生理学的研究も急速に進んだ。本稿では Marr の思想を背景に知覚と認知の共通の理論的基盤を求めて進めてきた我々の研究を概説する。

1　視知覚と視覚認知

視覚の機能は大きく二つに分けられる。まず視知覚機能である。視知覚は眼前にある面の構造および面の特性を二次元網膜像から推定する機能である。面の動きや面の色などの属性の推定も含む。したがって視知覚機能の本質は対象個別の知識をそれほど考慮せずに議論することができる。もう一つの機能は視覚認知機能である。これは視覚情報から対象物体が何であるのかを判断する機能である。何であるかを判断するだけでなく対象がどのように使うものか

といった機能をも判断しなければならない。たとえば、ハサミやコップを見たときにはそれが手の運動と結びついた形で、道具としての機能まで判断しなければならない（乾、1995）。これが視覚認知機能である。したがって視覚認知機能には対象個別の知識が重要な役割を果たす。

2　標準正則化理論とベイズ推定

一九八〇年代、視覚研究を中心とした脳研究で、Marr(1982) の思想は重要な役割を演じた。彼は、視覚の主要な役割が、「網膜に投影された二次元画像から外界の三次元構造を推測する」ことであると考えた。これは、数学的には不良設定問題（ill-posed problem）なので一般的には解くことはできないが、自然界の物理法則を問題の制約条件として用いることにより、脳は唯一の解に到達できると考えた。つまり二次元のデータしかないのに三次元の問題であるから明らかに未知数の数がデータの数より多い。しかし三次元解が満たす性質さえ分かれば解けるであろうというのが彼の考えなのである。その性質つまり問題の制約条件は個々の対象に依存しない物理法則である。この考えは、数学の逆問題を解く一般的な解法であるチーホノフの標準正則化法と全く同じ思想であることがその後指摘され、視覚の様々な計算が明確に定式化された（Poggio ら、1985）。

一般に、視覚の問題は $Az=y$ において「データ y（網膜像）」から「関数 z（推定された三次元構造）」を求めるいわゆる逆問題になっている。これは、一般に不良設定問題である。標準正則化理論では、

$$H=\|Az-y\|^2+\lambda\|Pz\|^2$$

なる評価関数を最小にする関数 z を見つける。$\|Pz\|^2$ は安定化汎関数で、関数 z に対する制約条件に相当するものである。λ は拘束条件の強さを決めるパラメータである。たとえば z が面の奥行きであれば制約条件は「面はほとんど

いたる所なめらかに変化しているであろう」ということになる（所々ある奥行きの不連続を除いてである。この理論を一般化すれば不連続をも見いだすことができるがここでは立ち入らないことにする）。これは

（評価関数）＝（データ回帰）＋（事前条件）

という形をしている。データ回帰項は、網膜像から直接計算されるデータとできるだけ矛盾しない面（関数）を選ぼうとするものである。くりかえすが、三次元の構造を二次元像から求めるわけであるから一般には解けない不良設定問題である。そこで外界の面に関する何らかの一般的に成り立つであろう物理法則を問題の制約条件としこれを満たすような面を求めるのが制約条件項である。

また標準正則化理論はある仮定の下ではベイズ推定を行うことと等価である。すなわち網膜像から三次元構造の推定をベイズ推定で行うのと等価である（たとえば、乾、1992b, 1993）。ベイズの定理を用いるとこれは次のように書くことができる。

$$P(\text{構造}\mid\text{画像}) \propto P(\text{画像}\mid\text{構造})\cdot P(\text{構造})$$

ここで、事前確率P（構造）は外界を支配する物理法則に対応し、条件付き確率P（画像｜構造）はある事象が生起したときに脳内である活性化の状態が生ずる確率を意味する。右辺第1項は光学系（広い意味で撮像系だけでなく神経の前処理も含めてもかまわない）の特性を、第2項は純粋に自然界の物理法則を表わしている。この意味で第1項をセンサモデルとか画像生成モデルとか一般化された光学と言うことがある。多くの可能性の中からベイズの定理を用いて事後確率を最大にする構造を眼前の構造であると推定することを最大事後確率推定（Maximum a posteriori estimate; MAP 推定）と言う。事前確率が与えられた場合、ベイズの定理から事後確率が計算でき、これを最大にする

事象を脳内活動の原因である眼前の構造とみなすのである。

個々の属性値がごく近傍の属性値にのみ依存するとき、画像がマルコフ性を満足するという。属性は、濃淡値、色、奥行きなどなんでもかまわない。通常網膜像はこの条件を満たしていると考えられる。物体の凝集性によってこのマルコフ性は成立するのである。このとき MAP 推定は、結局以下のような形のエネルギーの最小化に相当することが証明されるのである。すなわち、

MAP 推定＝U（画像｜構造）＋U（構造）の最小化

となる。U（画像｜構造）はある構造が生起したときにある画像が生ずる確率を決めるエネルギー、U（構造）はある構造が生起する確率を決めるエネルギーを表わす。条件付き確率はデータ回帰に、事前確率は制約条件に対応するのである。（マルコフ確率場理論については［乾 1992 c, d, e］参照。）これを確率的に解かず決定論的に解けば前述の標準正則化と同じことになる。

3　一般化画像放射照度方程式

網膜上に与えられる二次元画像データの生成過程（すなわち外界）をモデル化するとき様々のレベルでの記述が可能である。低いレベルでは、可視表面の奥行きや面の方向、各位置での反射率、照明光の強さと位置を決めれば画像データが決まる。より高いレベルでも視覚世界を記述できる。三次元空間の中に、別々の物体がどのように空間的に配置され、個々の三次元像は何で、それぞれがどのような並進・回転速度を持つかを記述しても、画像データを決定できる。脳内では、これらの様々の階層での記述がすべて使われていると考えられる。それを次の一般化画像放射照度方程式で表現できる（川人、乾、1990）。

$$I(\mu, x, y, \lambda, t)$$
$$=R(\nabla^2G*I, dI, d^2I, v, sd, r(\lambda), L, md, \nu, C, A, V, N, O)$$
$$=R(s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8, s_9, s_{10}, s_{11}, s_{12}, s_{13}, s_{14})$$
$$=R(S)$$

左辺は、左（$\mu=0$）か右（$\mu=1$）の網膜上の位置（x, y）での、時間t、波長λの光強度を示す。右辺は視覚世界の様子Sから画像データIが決まる画像生成過程を非線形関数Rで表わしたものである。すなわちRは、一般化された光学と呼ぶことができる。Rの中の引数 s_1〜s_{14} はすべて視覚大脳皮質で別々に表現され再構成されている。∇^2G*I は光強度と ∇^2G 関数の重畳積である。dI と d^2I はそれぞれ画像強度Iのある方向への1階微分と2階微分である。vは画像の濃淡値の最大変化方向の局所的な速度成分である。sd はステレオ視によって得られた奥行きを表わしている。$r(\lambda)$ は可視表面の波長λの光に対する反射率を表わしている。これはもちろん可視表面の場所に依存している。Lは観察者から見た可視表面の遮蔽輪郭や異なる物体の接合部などの不連続を表わす。md は単眼視によって得られた可視表面の奥行きを表わしている。νは照明光の波長分布と光源位置を表わす。Cは、Lで区別された個々の三次元物体の三次元的空間位置を示す。Aは個々の物体の色やテクスチャーなどの属性を表わす。Vは個々の物体の並進・回転の速度ベクトルを表わす。Nは観察者の身体や頭部・眼球が持つ並進・回転の速度ベクトルである。Oは三次元物体の記憶像を表わす。sd, $r(\lambda)$, L が 2・1/2 次元スケッチ（第5章参照）である。V, N, O を推定するのが高次視覚である。初期視覚、中期視覚、高次視覚かを問わず、IからSを推定する過程が視覚といえる。まとめると、

- Rは構造から画像への変換関数なので一般化光学と呼ぶことにする。
- 視覚はこの逆つまりIからSを解かねばならない。

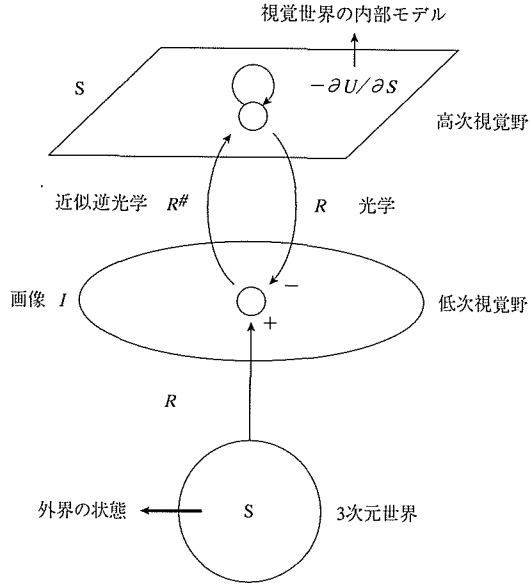


図1

4 視覚大脳皮質の計算理論

一般に視覚の働きは、逆問題を解くことであると述べた。もし線形の話であれば、それは逆行列あるいは疑似逆行列を求めればよいと言うことになる。問題によってはそれでよい場合もある。しかし一般にはそう単純には行かない。各モジュールの計算である逆関数は一般に求めることは難しい。従来のコンピュータビジョンの研究では、ある特殊な条件で成り立つようなアルゴリズムが多数考えられてきた（たとえば［乾 1992a］参照）。一方我々の知覚は非常に正確であることが多く、かつそれが短時間で計算される。脳のモデルであるためには、計算時間がほんの数百ミリ秒でなければならない。我々は、これが次のようにして実現されていると考えた（図1：川人と乾、1990）。

大脳には多くの視覚領野（モジュール）があり、それぞれの領野では視覚の異なる属性が処理されているらしい。これらの領野がいわば並列階層構造を

なしている。また視覚領野間には双方向性の結合が存在する。川人と乾（1990）では、領野間の双方向性結合により、逆光学と光学の計算が両方行われているのではないかと考えた。図1はこのことを単純化して示したものである。ここで重要な点は逆光学は厳密に逆関数でなくてもよく、ある限られた条件で成り立つようなものでよいのである。前向き（feedforward）結合によって、まず近似的になんらかの構造が推定される。つぎに後向き（feedback）結合によって、それを確かめるわけである。それは構造から網膜像を推定するわけであるから、光学を計算していることになる。次の計算ではこの誤差が前向きの計算に入力される。このようにするとたとえ逆光学が近似的であっても、正しい解に収束することが期待されるわけである。この近似がよいほど速く正しい答えが求められる。しかし実際の視覚系では、信号が網膜像にまで戻っているわけではない。網膜像から物体の構造や位置関係を直ちに計算するのは難しいため階層的に易しいものからむずかしいものへと順に解く必要がある。このなかでこのような計算がなされていると考える。この計算によって前述のMAP推定ができる。

Sの内部モデルの確率をP(S)，Sが与えられたときのIの条件つき確率をP(I|S)で表わす。2章で述べたMAP推定と同様に、それぞれのエネルギーをU(S)，U(I|S)とする。MAP推定に従って、次の事後エネルギーを最小化するSが推定されている。

$$U(S|I)=U(I|S)+U(S)$$
$$=1/2[R^{\#}(I-R(S))]^2+U(S)$$

ここで$R^{\#}$は画像生成過程Rの近似的逆モデルである。初期視覚で良く知られているように、画像生成過程の逆は不良設定であるからR^{-1}は存在しない。しかし、その近似$R^{\#}$は考えられるし、コンピュータビジョンで提案されてきた多くの一撃アルゴリズムは$R^{\#}$の具体例とみなせる。図1で、二次元画像データIは視覚下位中枢に、視覚世界の

様子Sは視覚上位中枢に表現されている。このモデルは視覚下位中枢を折り返しにして鏡像対称になっている。上位から下位への逆方向神経結合は画像生成過程Rの順方向モデルを与えている。一方、下位から上位への順方向神経結合は、画像生成過程Rの逆R^{-1}の近似逆モデル$R^{\#}$を与えている。さらに上位中枢内の固有神経結合はSの内部モデルとして$-\partial U/\partial S$を与えている。

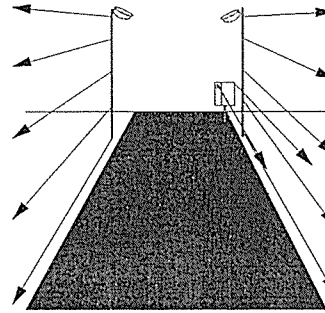
急速眼球運動などによって新しい画像データIが入力すると、下位から上位への順方向神経結合によってSの粗い推定値$R^{\#}(I)$が計算されるが、これはMAP推定にはなっていない。続いて、上位中枢の推定Sから逆方向結合によって、画像データの推定値R(S)が計算され、それが下位中枢で実際のデータと比較されて誤差$I-R(S)$が求められる。この誤差が順方向結合を通して上位中枢に戻されて、$R^{\#}[I-R(S)]$が入力される。一方、上位中枢内の固有神経結合は前式の第2項を与えている。この繰り返し演算によって、入力画像データを良く説明し、また内部モデルに照らして確率の高い、視覚世界の推定値が安定平衡状態として求められるのである。まとめると、

●脳は双方向の結合を利用して、順変換と逆変換を繰り返し、外界の構造を推定しているのではないだろうか。

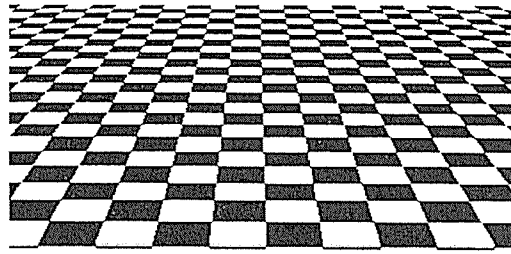
すでに述べたように視知覚機能の本質は、網膜像から眼前にある面の三次元特性を推定することである。網膜像には推定に必要なさまざまな手がかりが含まれている。脳内では、多くの処理過程が並列独立に網膜像に作用し、可視表面の幾何学的構造の(網膜座標での)表現を引き出す。これには、両眼視差(binocular disparity)、運動視差(motion parallax)、陰影(shading)、遮蔽輪郭(occluding contour)、テクスチャー(texture)などの処理が含まれる。これらの手がかりの一部について図を使って説明する。

図2a)は両眼視差の説明図である。いま、両眼である点Xを固視しているとする。このとき、Xの像は網膜の最も感度のよい中心窩と呼ばれる場所に投影される。両眼のレンズの結節点とXを通る円周上にある点は両眼の対応する(同一の)場所に投影される。しかし、この円の外側や内側の点は両眼でずれた場所に投影される。このずれを両眼

a）両眼視差

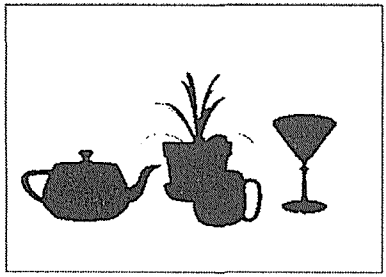


b）オプティカルフロー

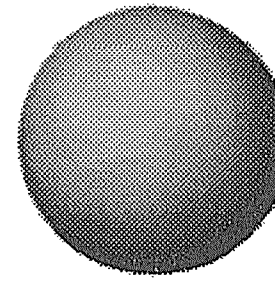


c）テクスチャー

図2



e) 遮蔽輪郭



d) 陰影

図2

視差と呼び、これを測定すれば相対的奥行きが求まる。ひらたく言えば、固視面を境に、それより遠くても近くても両眼の網膜像は図2a)のようにずれており、このずれを測定すれば奥行きが得られるのである。

図2b)は運動視差の説明図である。この図では対象が静止し、観察者が前進した場合のオプティカルフローが示されている。ベクトルの長さから相対的奥行きが求まる。もし、ある点に対してまっすぐに観察者が前進すれば、その点を中心に放射状のオプティカルフローが得られる。

図2c)はテクスチャーの勾配を表わす。この場合、無限にひろがる水平面が知覚される。図2d)は陰影の手がかりを表わす。この場合、球面が知覚される。画像の濃淡に対して局所的な微分操作を加えることによって、面の主曲率が計算できる。この原理を応用したものが化粧である。図2e)は図2a)で示した図の遮蔽輪郭である。

大脳には20から30の視覚情報処理に関与している領野がありそれぞれ名前が付けられている。図3にはこれまで知られている生理学や解剖学の知見をもとにして領野間の相互作用と各領野における情報表現の仮説が書かれている。このモデルでは、V2内の相互作用やV3、V4、さらにMTとV4のような色・形・動きの三つの左右の流れを統合する並列の相互作用もあるが、主な相互作用は左右の流れの中の上下の領野間の相互作用である。図3には、このような並列階層構造が多数含まれているがこれらの相互作用によって上記の逆計算がなされて各属

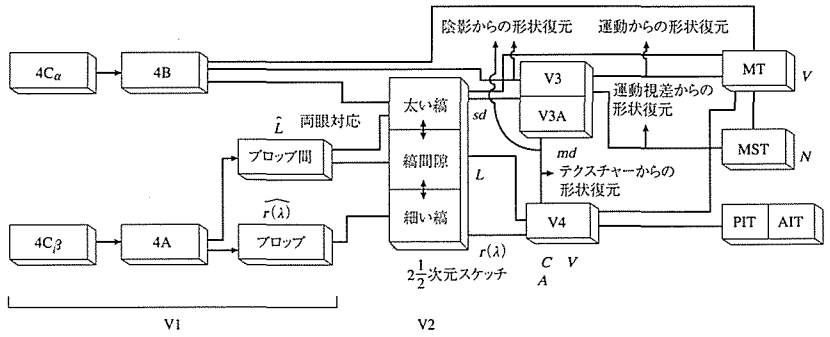


図3

性が推定されていると考えられる（川人、乾、1990；乾、1993）。

5　モジュールの統合

すでに述べたように網膜像から面の形状を推定する際、外界には多くの手がかりが存在する。たとえば両眼視差、動き、陰影、テクスチャーなどがあり、これらが別個のモジュールで処理されていると考えられている。多くのモジュールが存在し、それらが前述のように網膜像から面の構造を推定する計算を行っているということである。各モジュールの計算理論（すなわちこれらの手がかりのうちのひとつを用いていかに三次元構造が推定できるかという問題に関する計算理論）は、shape-from-X と呼ばれている。たとえば両眼視差であれば shape-from-binocular disparity、陰影であれば shape-from-shading である。その他Xには、texture, motion, occluding contour, surface contour などが入る。これらの問題は基本的には標準正則化理論によって説明されている。

したがって低次の視覚系（early vision, 初期視覚という）では、上で述べた多くの視覚情報に基づいて表面の幾何学的構造の推測を並列独立に行う。中間視覚（middle vision）では、これらのモジュールの出力を統合し、視線の方向によらない安定した表面の方向と奥行きを決める必要がある。表面の方向と奥行きの安定した記述を 2 1/2 次元スケッチと呼ぶ。もう少し正確

に言えば、2 1/2次元スケッチとは観察者中心の面（すなわち今見えている面だけ）の構造記述（奥行き、方向およびそれらが不連続に変化する位置の記述）である。もし2 1/2 次元スケッチが脳内で表現されているなら、中間視覚（middle vision）では、

1　複数のモジュールの出力情報がどのように統合され、一つの表現――2 1/2 次元スケッチ――になるのか。

2　より低次の段階では網膜中心座標で情報が表現されているので、どのようにして観察者中心すなわち頭部座標または身体座標に変換されるのか。

が問題となる。

モジュールの出力の統合に関してはさまざまな研究が進められている（Inui, 1996）。心理物理学的には単眼手がかりや両眼手がかりを組み合わせた時、形状がどのように知覚されるかが調べられてきた。その多くは、手がかりが矛盾したときに知覚される形状に関心があり、手がかり矛盾パラダイム（cue conflict paradigm）と呼ばれている。これらの実験結果から相互作用として次の二種類が考えられる。

(1)ある系の出力が完全に無視されてもう一つの系の出力のみによって計算される。

(2)複数個の出力があればそれらの重みづけ平均によって最終的な知覚が決定される。

乾と山下（1993）はさまざまな心理物理実験の結果から、脳はこの情報統合もベイズ推定に基づいて行っているのではないかという仮説を提案している。この場合のベイズ推定は複数の情報源から外界の構造を推定するということに対応している。モジュールの出力が加算平均されるという結果は、ベイズの定理からすれば、複数のセンサーを用いて泥棒が入ったという仮説を検定する問題とちょうど等価になっている。Sを仮説、e_k は異なる種類のセンサーの信号とし、n個の独立に働くセンサーがあるとする。いくつかの異なる情報で泥棒を検出すると考える。すなわち、音の変化や温度の変化やあるいは気流のようなものである。この時条件付き確率は、

$$p(e_1, e_2, \cdots, e_n|S) = \prod_{k=1}^{n} p(e_k|S)$$

となり，事後確率は

$$p(S|e_1, e_2, \cdots, e_n) \propto P(S) \prod_{k=1}^{n} p(e_k|S)$$

と書ける。このとき確率分布に Gibbs 分布を仮定すれば，条件付き確率の積は評価関数のデータ回帰項の加算に対応する。すなわち，センサーが二種類のとき評価関数のデータ回帰項は，

$$\lambda_1 \sum (f_i - g_i)^2 + \lambda_2 \sum (f_i - h_i)^2$$

となる。ここで f_i は未知関数 g_i と h_i は位置 i における異なるモジュールの出力データである。したがって，面の知覚が両眼立体視系の出力と単眼立体視系の出力の加算平均になっているという事実はベイズ推定の枠組みからは二種類の検出器が並列独立に働いて外界の構造を推定していることと等価になる。

もう少し一般的に書くと，

$E = \lambda_1$（検出器1によるデータ回帰）
$\quad + \lambda_2$（検出器2によるデータ回帰）
$\quad + \lambda_3$（制約条件）

ここに

λ_1, λ_2：データ回帰の制御パラメータ（信頼性，優先度に依存）

である。なお、朝倉と乾（1994）は、推定される面の曲率すなわち $\partial^2 f/\partial x^2$ に従って制約条件に対する重みを変化させることにより、両眼立体視に関するいくつかの現象のシミュレーションを行っている。さらにわれわれはテクスチャーを構成する個々の要素であるテクセルの密度情報とテクセルの扁平率の手がかりがどのように統合されて面の知覚が形成されるかを心理物理学的に検討している。梅村と乾（1995）はこのような実験を行うことによって二つの手がかり間の統合がやはり標準正則化理論に基づいて正確にシミュレーションできることを示した。これは別の言い方をすればベイズ推定によって面を知覚していることを示すものである。

6 情報統合の実現（implementation）に関する示唆

八〇年代、高次視覚野の解剖学的、生理学的研究が急速に進みそれらの構造と機能が部分的に解明されてきた（乾、1993 参照）。視覚野は20～30の領野に分けられ、それらの間の結合も明らかにされつつある。重要な特徴は、

1 情報処理が完全な並列構造でもなく単純な階層構造でもない。いわば並列階層構造をしている（図3）。

2 ある種の属性を処理するモジュールが存在し、それが階層的に処理が進められる。例えば、色情報は V1 の blob→V2 の thin stripe→V4 といった具合いである。

3 領野間は、双方向に結合している。area A から area B へ信号を伝える経路があれば必ず逆に area B から area A への直接的経路が存在する。

4章で述べたように、川人と乾（1990）は、双方向性結合に注目し、このような結合を通じて視覚計算が速くかつ正確に進められることを理論的に示している。この理論では近似逆光学を前向き結合で光学を後ろ向き結合で行う。大ざっぱに言えば、ボトムアップで仮説生成、トップダウンで仮説検証をこのサイクルを通して速く正確な答えを求

めるというものである。また Inui(1992) は、双方向性結合の可能性のある機能として

(i)　近似逆光学と光学
(ii)　モジュール間の首尾一貫性保持
(iii)　後ろ向き結合による選択的注意

を挙げている。

また、Zeki らは backward 結合が直接の forward 結合がない領野へも存在することから、これによって情報の統合がなされているのではないかという re-entry 仮説を提案している。例えば、V2-thin stripe (color & form)→V4 および V2-thick stripe (stereo & motion)→V5 (also called as MT) という前向き結合では明確に信号が分離しているにもかかわらず、後向き結合では例えば V5→V2 というように V2 の thick stripe に信号をもどすだけではなく thin stripe にも信号をもどす。彼はこれにより形と動きの情報が統合されるのではないかと考えている (Zeki and Shipp, 1989a,b)。

7　パターン認識の計算論的枠組みと知識の統合

網膜情報は外側膝状体を経て後頭部にある一次視覚野に到達する。一次視覚野から視覚情報は二つに分かれる。一つは頭頂連合野に伝達され、もう一つは側頭葉下部に伝達される。下側頭葉 (IT) は、視覚パターン認識および記憶の中枢であると考えられている。ここにおいて視覚情報は、パターン認識のために最適な表現形式に変換されているはずである。しかし Marr (1982) の三つの水準のいずれについてもほとんどわかっていないのが現状である。

おそらく、脳内の表現空間は対象間の類似の度合いに応じた位置関係を保存する位相空間になっているはずである。一方、画像から様々な特徴量を計測するとそれらの間に相関が見られることが多い。したがって、パターン認識にお

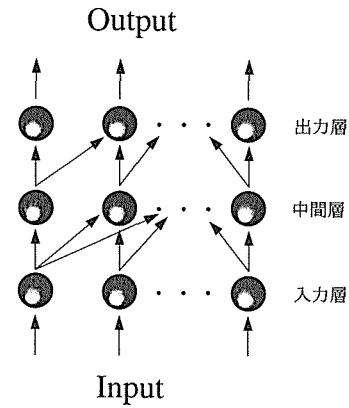


図4

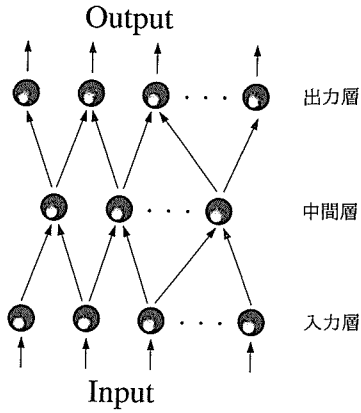


図5

いては、特徴空間の次元削減が重要になる。以上の事実から、人間の類似性のデータと主成分空間との比較を行うことは重要であると考えられる。

形の内部表現の解明を目的として、我々は、人間が感ずるパターン間の類似度に着目した。心理学では、類似度を直接評定する方法と瞬間提示条件などでの混同率を測定する方法がある。混同率を分析する場合は、混同率が高いほど類似度が高いと考える。そして類似度が高いほど脳内の表象空間において近い距離で表現されていると考えるのである。被験者のバイアス等を考慮した上で多次元尺度構成法を用いると空間上での配置を視覚的に表現することができる。これまでも心理学的にパターン間の類似度を調べた研究はあった。しかしいずれも結果の解釈は使われたパターンに強く依存したものであった。一般原理を探るためにはどうあるべきかを深く考察する必要がある。そこで次に人工ニューラルネットワークを用いて以下のような研究を行った。

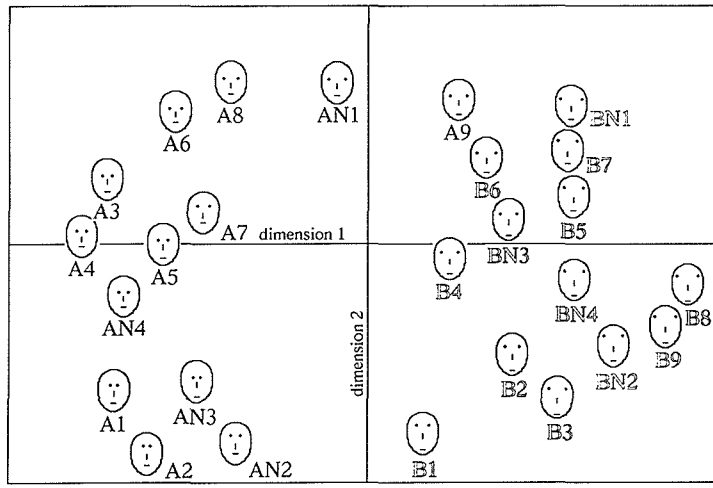
ニューラルネットワークにおいては、個々のニューロンは単純な演算を行っており、それぞれシナプスで結合している（図4）。図では、黒丸でニューロンの細胞体を、矢印で軸索を表わしている。球と矢印の接合部をシナプスという。学習等によりこのシナプスの

伝達効率すなわちニューロン間の結合の強さが変化する。ニューラルネットワークでは、多くのニューロン間の結合係数（重み）によって情報が表現されている。したがっておのおのの情報は一つの処理ユニットによって表現されるような局所表現ではなく、処理ユニット全体の活動パターンとして表現される。これを分散表現と呼ぶ。分散表現を用いると多くの情報が同じシステムに多重に記憶させられるだけでなく、さまざまな興味ある現象が説明できる。並列分散システムによる学習では、新しい情報をメッセージとして獲得するのではなく、受け取るデータを最小の相違で適合するように自分自身を調整するのである。つまりシステム自身が変化するのである。学習アルゴリズムは、ネットワークの出力が望ましい出力に近くなるようにシステム内部の重みを変更するものである。ここで用いたニューラルネットの構造を図5に示す。ニューラルネットは三層構造をしており、入力層、中間層、出力層からなる。図4と異なる点は、中間層のニューロンの数が入出力層に比べて少ないことである。入力として心理実験で用いられたパターンと同じパターンのセットをこのニューラルネットに提示し、入力と同じ大きさの出力が出力層から出るように学習させる。このとき中間層の細胞の数を入出力層よりかなり少なくしておくのである。このネットワークはこのような構造から砂時計型ニューラルネットと呼ばれる。つまり砂時計型ニューラルネットワークは、中間層に圧縮された表現を作るように結合荷重を調整していくのである。うまく学習が進むとネットワークは、情報の圧縮法を獲得できるだけでなく、圧縮された表現からもとの入力を復元する方法も獲得するのである。入力と出力が等しい変換を学習するので恒等写像の学習と言われる。重要な点は、

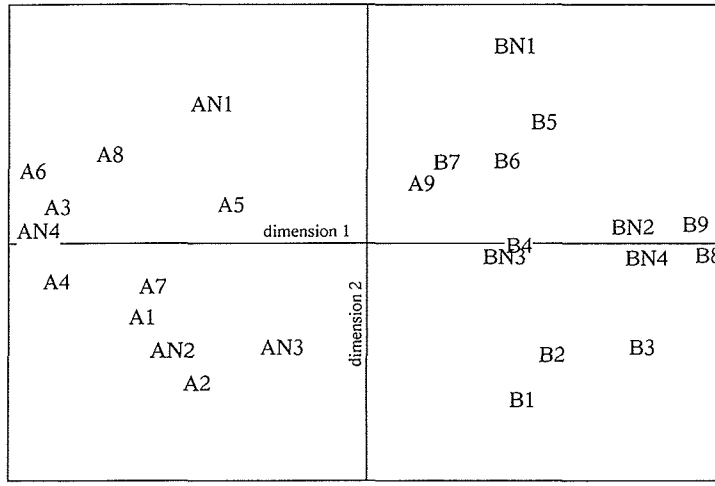
●恒等写像を学習するニューラルネットの中間層には刺激パターンの圧縮された表現ができる。

●それは、一個のニューロンが一個のパターンを表現しているのではなく、中間層にあるニューロン（シナプス）全体で表現されている（分散表現）。

ということである。



a) 顔図形の類似性空間（吉川，1980 より）



b) ネットワークの内部空間

図 6

心理実験で用いられた刺激セットを三層の砂時計型のニューラルネットワークに学習させ、その内部表現を分析した。学習後ネットワークの内部表現を調べたところ、被験者が感ずる類似性空間とほぼ一致する空間を構成していることが分かったのである（図6）。図6aは顔図形に対して人間が感じた類似度を表現しており、顔図形の距離が類似度に対応するように描かれている（吉川、1998）。図6bは、ニューラルネットワークを用いて得られた圧縮表現の距離を同様にして描いたものである。両者の類似性の関係はきわめてよく似ている。恒等写像を学習する三層のニューラルネットワークが主成分分析に近い働きをしていることが理論的、実験的に示されている。したがってこの結果は、人間の持つパターンの内部表現が、主成分分析に近い性質を持っていることを示している。より正確に言えば、各パターンに対するハイパーコラムの出力の主成分分析の結果が人間の類似度空間にきわめて近いのである。同様の結果は、顔パターンのみならず、文字、ドットパターン、幾何学図形などを用いても確認されている（たとえば、牧岡ら、1993, 1994; Makioka ら、1996; 森崎、乾、1995, 1996）。

これは何を意味しているのだろうか。まず、パターンの類似度は与えられたパターン集合に依存するということである。つまり、二つのパターンの類似度は過去にどんなパターンを見たかによって異なる。おそらく、パターンを見るごとにその表現がパターン空間でなるべく距離が長くなるように修正されているのであろう。学習によって継時的に獲得される知識の統合過程を示唆していると思われる。一方、Young と Yamane (1992) は、サルの IT において顔パターンがどのように表現されているかを調べた。その結果、顔パターンが多数のニューロン集団の活動パターンによって符号化されていること、ならびに活動パターンによって、顔の類似度なども表現されているらしいことなどを発見している。我々の研究は、活動パターンを構成する個々のニューロンがどのような原理でパターンの符号化をしているかについての仮説をも提案するものである。実際、森崎と乾 (1996) は、砂時計型のニューラルネットの中間層のニューロンの反応選択性を調べたところ Young と Yamane (1992) が IT で記録したニューロンの反応選択性と

きわめてよく一致していた。また Inui, Morisaki と Sugio (1996) は、いくつかの研究で公刊された IT ニューロンの反応選択性を砂時計型ニューラルネットでシミュレーションできることを示している。

ところでもう一度図5を見てみよう。砂時計型ニューラルネットでは、恒等写像を学習するのであるから、図7のように折り返して中間層の出力を入力層にフィードバックして、入力とフィードバック信号の誤差を小さくするように学習を進めればよい。ネットワークの機能はいずれでも同じなのだが、図のようにすると知覚の章で論じた双方向性結合による情報処理（図1）との関連性が出てくる。この場合は、前向きで情報圧縮を後ろ向きで信号の復元を行うのである。また6章で述べたように、Inui (1992) は、双方向性結合の可能性のある機能として

(i) 近似逆光学と光学

(ii) モジュール間の首尾一貫性保持

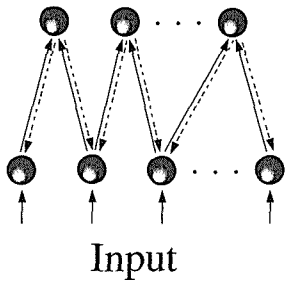


図7

(iii) 後ろ向き結合による選択的注意

を挙げているが、これに加えて

(iv) データ圧縮と復元

の機能を追加しなければならない。iとiiは、主として知覚レベルの機能である。

Barlow と Földiák (1989) によって提案されたアンチヘッブ anti-Hebb という学習則もデータ圧縮と関連している。いずれの場合も、データ圧縮とはデータに含まれる相関を出来るだけなくし無相関なベクトルに変換することにより、冗長度を少なくすることである。上述のように我々は心理学的に得られた文字や幾何学図形、顔の心理的類似度が主成分分析（KL 変換）と密接な関係にあ

ることを見いだした。データ圧縮と復元という枠組みによって人間のパターン認識機能も体系づけられるのではないだろうか。

8　情報統合の理論構築を目ざして

言うまでもなく、われわれは脳の中で作り上げた環境のモデルの中で行動している。モデルは経験によって作られ、修正されていく。われわれは、このようにして作り上げた環境のモデルに基づいて対象を理解している。では情報の統合はなぜ必要なのだろうか。もちろん、様々な状況下で適切な行動を行うためである。具体的にその目的を二つに分けることができる。まず過去の経験に基づき、データの信頼性を正しく判断し、正しく外界の構造を推定することである。様々な情報源から推定される構造の一貫性を維持することも必要である。本稿では、基本的には、ベイズ推定の立場に立ちそれを評価関数最小化という方法で実現できることを述べた。これは、外界のモデルをベースにした方法である。

一方適切な行動をするためには、外界のモデルを正しく構成し、正しくそれを利用することが必要である。観察者自らの状態によらず外界の物理的特性を正しく評価するためには、不変な物理的属性をうまく推定することが必要である。知覚の恒常性（視空間の安定性、対象の物理的大きさの評価など）はその典型例である。恒常性を得るためには、異種情報を統合しなければならない。前述の外界の構造推定においても外界のモデルである拘束条件（モデル）の信頼性をうまく取り込まねばならない。外界の知識をうまく獲得するために人間は、主成分分析に近い統計的手法を用いて位相空間を脳内に作り上げていることも述べた。この手法は、分散最大化という方法で情報縮約を行う。うまく神経回路網で学習すると正しく原パターンを再生することもできる。我々は現在、知覚と認知の問題に共通した理論的基盤の構築に向けて研究を進めている。

文献

朝倉暢彦，乾　敏郎（1994）両眼視差による面復元過程のニューラルネットモデル　電子情報通信学会技術研究報告，NC94-21, 55-62.

Barlow, H. B. and Földiák, P. (1989) Adaptation and decorrelation in the cortex, In: Durbin, R., Miall, C. and Mitchison, G. (Eds.) The Computing Neuron, Addison-Wesley Publishing Company.

乾　敏郎（1992a）3次元世界の再構成　認知科学ハンドブック　共立出版

乾　敏郎（1992b）知覚の認知科学――脳における視覚計算を考える――　人工知能学会誌，7, 15-23.

乾　敏郎（1992c）確率的画像処理とニューラルネット1。やさしいマルコフ確率場入門　画像ラボ，第3巻，5号

乾　敏郎（1992d）確率的画像処理とニューラルネット2。やさしいマルコフ確率場入門　画像ラボ，第3巻，6号

乾　敏郎（1992e）確率的画像処理とニューラルネット3。平均場近似とボップフィールドネット　画像ラボ，第3巻，7号

Inui, T. (1992) Computational considerations on the possible functions of the backward connections between brain modules. Proceedings of the 1st Asian Conference in Psychology.

乾　敏郎（1993）Q & A でわかる脳と視覚　サイエンス社

乾　敏郎（1995）知覚と運動　乾編著「知覚と運動」（東大出版会）

Inui, T.(1996) Mechanisms of information integration in the brain. In: T. Inui, and J.L. McClelland(Eds.) Attention and Performance XVI. The MIT Press.

乾　敏郎，山下博志（1993）中間視覚における情報統合のメカニズム　電子情報通信学会技術研究報告，NC93-13, 9-16.

Inui, T., Morisaki, A., and Sugio, T. (1996) How does the brain represent and categorize visual shapes? Proceedings of the 3rd conference of Brain and Mind IIAS Japan.

川人光男，乾　敏郎（1990）視覚大脳皮質の計算理論　電子情報通信学会論文誌，J73-D-II, 1111-1121.

牧岡省吾，乾　敏郎，山下博志（1993）文字パターンの心理空間と脳内表現　ニューロコンピューティング研究会資料，NC93-35, 33-40.

牧岡省吾，乾　敏郎，山下博志（1994）2次元パターンの脳内表現――ハイパーコラム的入力表現を用いた検討――　日本心理学会第58回大会発表論文集

Makioka, S., Inui, T., and Yamashita, H. (1996) Internal representation of two-dimensional shape. Perception, 25. (in press)

Marr, D.(1982) VISION-A Computational Investigation into the Human Representation and Processing of Visual Information, W. H. Freeman and Company. 乾　敏郎，安藤広志（訳）1987 ビジョン，視覚の計算理論と脳内表現　産業図書

Poggio, T., Torre, V., & Koch, C. (1985) Computational vision and regularization theory, Nature, 317, 314-319.

梅村浩之，乾　敏郎（1995）形状知覚に及ぼすテクセルの密度および扁平率の勾配の効果　電子情報通信学会技術研究報告，PRU 95-85，HIP95-13, 37-42.

吉川左紀子（1980）類似性構造に基づく図形分類反応の検討　心理学研究，51, 267-274.

Young, M. P. and Yamane, S.(1992) Sparse population coding of faces in the inferotemporal cortex, Science, 256, 1327-1331.

Zeki, S. and Shipp, S.(1989a) The organization of connection between areas V5 and V2 in macaque monkey visual cortex. European Journal of Neuroscience, 1, 333-354.

Zeki, S. and Shipp, S.(1989b) Modular connections between areas V2 and V4 of macaque monkey visual cortex. European Journal of Neuroscience, 1, 494-506.

（筆者　いぬい・としお　京都大学大学院文学研究科〔心理学〕教授）

Computational Theory of Visual Perception and Cognition

by Toshio INUI
Professor of Psychology
Graduate School of Letters
Kyoto University

Marr's philosophy has played a significant role in studies of the brain, notably in the vision studies during the 1980s (Marr, 1982). He proposed that the major function of vision is to estimate the 3-dimensional structure of the world from a 2-dimensional image projected onto the retina. Mathematically, this is an ill-posed problem and a general solution cannot be given in most cases. He suggested that a unique solution to this ill-posed problem could be given if physical laws were taken into consideration as constraints. It was later pointed out that this idea of Marr's was conceptually equivalent to Tikhonov's method of standard regularization, which is a common method for solving inverse problems in mathematics (Poggio, et al., 1985). With this realization, various visual computations have been formulated in precise mathematical terms.

In addition, Marr suggested that there are many modules in early vision, and that the computation for estimating of surfaces from the retinal image is carried out independently in each module. The computational theory of modules is now generally called "Shape-from-X", and it has generated a large body of research. Here, X represents binocular disparity, shading, texture, motion, or some other source of information relevant to shape determination.

In this paper, we initially discuss how outputs of early vision modules are integrated into one unique representation: a 2 1/2D sketch. For this problem, a Bayesian estimation framework is useful for explaining much of

the psychophysical data obtained by a cue-conflict paradigm. We proposed a new theory for the integration between vision modules that is based on a Bayesian estimation and a simple neural network.

On the other hand, a representational space should be a topological space that preserves a relative degree of similarity between patterns. In order to investigate the representation of a visual pattern, it is very important to develop a general framework that explains their pychological similarity. In the latter part of this paper, the mechanism of visual cognition is discussed based on our recent research concerning psychological similarity. We then introduce a neural network model of the inferotemporal cortex, which is the center of visual cognition.

Finally, we discuss the computation of information integration in middle vision and pattern representation in the general framework of visual computation proposed by Kawato and Inui (1990).

Magna Propensio chez Descartes

——Sur la preuve de l'existence des choses corporelles——

par Takashi KURATA
professeur adjoint
à l'Université de Shimane,
Faculté de droit et des lettres,
Philosophie

Dans la VI[e] Méditation, concernant l'existence des choses corporelles, Descartes dit ainsi: «Dieu m'a donné une très grande inclination (*magna propensio*) à croire que les idées des choses sensibles me sont envoyées des choses corporelles.» Il prouve l'existence des choses corporelles en s'appuyant sur cette inclination. Dans ce traité, en mettant au point cette «grande inclination», je voudrais examiner la pertinence de la preuve de l'existence des choses corporelles que montre Descartes dans sa VI[e] Méditation.