ニューロコンピューティングによる感覚情報処理

杉 江 昇*

1. はじめに

人間の視聴覚等の感覚情報処理能力についての研究は、心理学、神経科学の分野で着実に進歩を続けている^{1)~3)}。また、理論的な分野においても、計算理論と呼ばれる枠組の提唱により、両眼立体視、運動立体視などの研究に大きな進展が見られる⁴⁾。さらに、これらの処理を神経回路網で実現する方法についても、いくつかの提案がされているが、このような提案は LSI によるハードウェア化へつながるものとして期待されている。以下では、視覚を中心にして、上述のことを中心に概観する。

2. 視覚系の概要^{1)~3)}

よく知られているように、視覚系のセンサ部は眼球の網膜である。網膜には視細胞と呼ばれるセンサがあり、これで外界各部の明るさや色の検出を行う。視細胞の出力は、中継細胞を介して、神経節細胞と呼ばれるニューロンへ送られる。神経節細胞には大型のものと小型のものがあり、その出力が、外側膝状体 (LGN) と呼ばれる部分のそれぞれ大型細胞層 (Magno) と小型細胞層のニューロンへ伝えられる。図1に示すように LGN の小型細胞層のニューロンは、脳の第1次視覚野 (V1) のブロッブ (Blob), インターブロッブ (Inter Blob) と呼ばれる部分のニューロンへ情報を伝える。他方大型細胞層のニューロンは、4B 層と呼ばれる部分のニューロンへ情報を伝える。脳には視覚に関連したモジュールは V1 だけでなく、V2, V3, V4, MT などの名を持つものが知られていて、それらのサブモジュールやモジュール相互の接続が次第に明らかになってきている。図で太い実線は強い主要な接続経路を、細い実線は弱い2次的な接続経路を示す。破線は十分に確認されていない接続経路を示す。接続は前向きだけでなく後向きもあるのが普通である。図で箱の中のアイコンはそのモジュール中のニューロンがどのような視覚刺激に反応するかを示す。

眼鏡は両眼刺激選択性、人差指は運動方向選択性、角度記号はエッジ方向選択性、虹記号は色選択性であることを示す。

以上から、視覚系には、いくつかの機能の異なるモジュールがあり、各モジュール間には階層的な関係があることなどが、推測できる。階層の一番上に示されているのは、顔などの識別に関与する下側頭野 (Inferotemporal Areas) と、対象物体の空間定位に関与する頭頂野 (Parietal Areas) とである。

3. 計算理論

感覚情報処理のモデルは、それまで、心理学的知見を、神経生理学的知見に若干の仮説を付加して説明するのが普通であった。これに対し、処理を説明するのに、計算

図1 赤毛ザル視覚系のモジュール構造¹⁾

原稿受付 1991年3月14日

Sensory Information Processing by Means of Neuro-computing

* 名古屋大学工学部

表 1 初期視覚における正則化⁵⁾

Problem	Regularization principle
Edge detection	$\int [(Sf-i)^2 + \lambda(f_{xx})^2] dx$
Optical flow (area based)	$\int [(i_x u + i_y v + i_z)^2 + \lambda(u_x^2 + u_y^2 + v_x^2 + v_y^2)] dx dy$
Optical flow (contour based)	$\int [(V \cdot N - V^N)^2 + \lambda((\partial/\partial_s) V)^2] ds$
Surface reconstruction	$\int [S \cdot f - d]^2 + \lambda(f_{xx}^2 + 2f_{xy}^2 + f_{yy}^2) dx dy$
Spatiotemporal approximation	$\int [(S \cdot f - i)^2 + \lambda(\nabla f \cdot V + f_t)^2] dx dy dt$
Colour	$\ I^* - A_z\ ^2 + \lambda\ P_z\ ^2$
Shape from shading	$\int [(E - R(f, g))^2 + \lambda(f_x^2 + f_y^2 + g_x^2 + g_y^2)] dx dy$
Stereo	$\int [\nabla^2 G * (L(x, y) - R(x + d(x, y), y))]^2 + \lambda(\nabla d)^2] dx dy$

理論のレベル、表現とアルゴリズムのレベル、ハードウェアによる実現のレベルに分けて論じようというのが計算理論の立場である⁴⁾。この立場からすると、従来のモデルは、ハードウェアによる実現のレベルのみを考えていたことになる。計算理論のレベルでは、コンピュータビジョンの理論的研究と、共通の土俵で抽象的理論のレベルで議論が行われることになる。

コンピュータビジョンの分野では、陰影からの形状復元や、オプティカルフロー（2次元速度場）を求める問題が、表面の滑かさや、速度場の滑かさを制約条件とする変分問題として定式化できることが知られていた。このような定式化が、生体の初期視覚でも成立つものと考えられている⁵⁾（表 1）。

上述の考え方に沿った研究の一例にバイオロジカルモーションのモデルがある。人体の各部に豆電球をつけ、

(a) 1本目のロッドの回転を打ち消した状態

(b) 1, 2本目のロッドの回転を打ち消した状態⁶⁾

図 3

図 2 3本のロッドの連結物体の運動⁶⁾

暗室で歩きまわらせ、それを映画に撮りスクリーンに投射すると、個々のコマには、十数個の光った点しか写っていないのに、映画になると人体が知覚される。これがバイオロジカルモーションの知覚である。そのモデルとして、豆電球が、さまざまな長さの目に見えないロッドに取り付けられ、各ロッドがそれぞれ固有の軸まわりに回転していて、ロッド同士が関節で結合され、結合により最低一つ以外は他のロッドの回転の影響を受ける物体を考える。そうすると、正射影速度場から、ロッドの3次元運動と構造を導くことができる。3枚の速度場画像から、2通りの解がえられる⁶⁾。図2は3本の連結ロッドの場合で、左のロッドは空間に対し固定した方向の軸まわりに回転しているが、まん中および右のロッドは、連結の影響で、回転軸は一定でない。しかし、まず左のロ

(a)

(b)

図 4 主観的輪郭を生じる図形⁷⁾

図 5 抽出された物理的輪郭（実線）と主観的輪郭（破線）⁷⁾

図 6

ッドの3次元の運動および構造が復元できることを発見でき、ついで、左のロッドの運動を打ち消すと、まん中のロッドの回転軸が一定となり復元が行え、ついで同様にして、右のロッドの3次元運動と構造が復元できる。

もう一つの例として、主観的輪郭を取りあげる。図 4 を見ると、白い三角形や長方形が、他の図形の上に置かれているように見える。これら白い図形の輪郭は、物理

図 7 時間差検出神経回路⁷⁾

図 8 音源とマイクロホンセットの配置⁸⁾

的には一部しか見えないが、主観的には確かに見える。このモデルでは、ある図形が他の図形を、おおい隠すと線が途切れたり、T形の頂点ができたりすることに着目する。白い図形と背景との間には明るさの差がないためT形の頂点はL形となる。また白い図形は、ほぼいたる所で滑かな輪郭を持つ。このようなことを考慮すると、図 5 に示すように、主観的輪郭をコンピュータで抽出することができる。主観的輪郭の生成機構は、図 6 (a) のような重なり合った図形の外輪郭から、各成分図形を推定するのにも役立つ。

両眼立体視に相当する聴覚機構は、両耳聴による音源の空間的位置の決定である。その手がかりの一つとして、両耳へ到達する音の時間差が用いられている。時間差を求めるには、図 7 のようなディレイライン付一致回路が脳では用いられている。時間差を取るといっても、両耳

の受ける音響信号の特徴的な部分—反響の影響を受けにくい立ち上がり部分—の間の対応づけを取ることが先決である。ここでは、時間差の絶対値がたかだか、両耳間を音が通過するのに要する時間以下であるとの制約が対応候補を絞るのに用いられる。時間差だけだと、距離差一定の点の軌跡（双曲線）上にしか、音源の位置を限定できない。受音器が無方向性であり、静止しているとする、図8のように、2つでなく3つ必要である。（受音器と音源が同一平面上にあるとして）このような設定で、たとえ反響のある部屋の中でも、2つの音源の存在方向

を数度の精度で決定できる。さらに、音源の方向が分かると、混合した2つの音をマイクで受けても、各音源ごとに分離することもできる。このような仕組みは、移動ロボットを声で呼び寄せたり、死角からの音を聴いて、そちらの方向へカメラを向けたり、接近したりして、発音物体を確認するのに使えるだろう。

図9 線の方角閾値と線の長さ、呈示時間⁹⁾

図10 刺激呈示時間と両眼視差閾値⁹⁾

図11 陰影と両眼視差の組合せによる奥行知覚¹²⁾

4. 多チャネル理論

心理物理実験と、神経生理学的実験とにより、視覚系は、空間周波数帯域の異なる、すくなくとも4つのチャネルから構成されていると考えられている。各チャネルの最初の処理は、 $\nabla^2 G(\sigma)$ 演算子でモデル化できる。この演算子は、無方向性のエッジ抽出演算子で、ガウスの誤差関数の σ が大きいと、低周波帯域フィルタ、小さいと高周波帯域フィルタとなる⁹⁾。

多チャネルの一つの存在理由として、両眼立体視でのエッジ間対応を、 σ 大のチャネルから順次 σ 小のそれへ適用することにより、効率よくかつ高精度で行えるのではないかと仮説が知られている⁹⁾。

これに関連して、線分の方角の判定閾値を調べた実験がある。図9がそれで、線分が長いほど閾値が小さいこと、呈示時間が長いほど閾値が小さいことが分る。理論的検討により、呈示時間が長くなるにつれ、処理系の σ が小さくなると考えられる。同様に、図10は、両眼視差の閾値が呈示時間と共に小さくなっていくこと、あるいは σ の大きなチャネルから小さなチャネルへと処理が移行していくことを示すものである。

以上のような仕組みの画像処理への応用として、テクスチャの領域分割が試みられている^{10,11)}。

5. 視覚モジュール間統合

図1に示したように、視覚系は、いくつかの機能別モジュールから構成されていて、各モジュールでの処理を統合することにより、効率のよい高信頼性の処理を実現しているように思われる。

図11は、陰影と両眼視差との統合により奥行き知覚がどう影響を受けるかを示す心理実験結果である。対象は回転楕円体で、その長軸と短軸の比がElongation, n は実験回数である。図で、陰影にもエッジにも視差のある刺激(D^+E^+)からの奥行き知覚は、ほぼ正確である。エッジが見えない D^+E^- では、奥行き知覚は70%位に減る。陰影のみで視差情報がまったく与えられない D^-E^- や D^-E^+ では、奥行き知覚はきわめて弱い¹²⁾。

図1や、図11の知見を取り入れたモデルを、64kのプロセッサからなる並列計算機 Connection Machine 上で実現したのが図12である。ここでは、色、テクスチャ、動き、両眼立体視、明るさのエッジのモジュールからの情報を統合し、物理的不連続部を抽出することが試みられている。

もう一つの例は、能動的なカメラの移動と組合せることにより、陰影からの形状復元とテクスチャからの形状復元を統合する試みである¹⁴⁾。

図12 視覚モジュールの統合¹³⁾

図13 (a) 水平(I_h^h)と垂直(I_h^v)のラインプロセス、(b) 不連続を考慮したオプティカルフロー計算のための回路網¹⁵⁾

6. ニューラルネットワークによる実現

半導体技術の発展により、感覚情報処理機構のハードウェア化が試みられている。以下にはその観点から、ニューラルネットワークモデルを紹介する。

オプティカルフローあるいは速度場 (u, v) については、表1のところですでに述べたように、画像の明るさの時間的变化 i_t 、空間的变化 i_x, i_y との間に、下の関係が成立ち、

$$i_x u + i_y v + i_t = 0$$

また、速度場の滑かさから

$$u_x^2 + u_y^2 + v_x^2 + v_y^2$$

をできるだけ小さくしたいとの要請がある。さらに異なった動きをする物体間で生じる速度場の不連続を扱うた

図 14 陰影から表面曲率を学習する神経回路網¹⁶⁾

図 15 (a) 直方体および (b) 三角柱の見え方の視点による変化¹⁸⁾

めに、ラインプロセスと呼ばれる0か1の値を取るものも含めた最適化が必要となる。このような仕組みは、図13に示すようなアナログの抵抗回路網で実現できる。図13(b)の回路網で上の層の格子点の電圧は、各点での u に、下の層のそれは v に対応する。寄生容量による動特性に従って落ち着いた状態がオプティカルフローの解となる。

第二の例は、陰影からの形状復元を行うニューラルネットワークである。第一の例は、計算理論から直接導か

れたネットワークであるが、第二の例は、教師付学習により導かれる。すなわち、図14に示すような多層パーセプトロンに、誤差逆伝搬法を適用して、層間の結合係数を修正すると、陰影から表面の主曲率をかなり正確に計算できるようになった。このようにして、計算理論を介さなくても、問題を解くニューラルネットワークを求めることができた。さらに、学習後のニューラルネットワークを調べると、ネコやサルの大脳視覚野で見出されている単純型ニューロンや、端点選択性ニューロンが形

図16 一つの(a)完全なあるいは(b)不完全な見え方を提示すると他の見え方を想起する¹⁸⁾

成されていた。このことは、単純型ニューロンなどが、2次元画像処理のためと考えられていたことに対し、3次元形状抽出に関与している可能性を示すものとして注目される。

第三の例は、3次元物体は、視点を変えると見え方が変ることに関するものである^{18,19)}。図15は視点によって変る見え方の系列である。このような系列を、連想記憶用ニューラルネットワークに記憶させると、直方体の一つの見え方を入力すれば、直方体の他の見え方が想起される。想起のための入力は、一つの見え方の一部でもよい。三角柱も直方体も同一のネットワークに一緒に記憶させることができるので、三角柱を見せれば三角柱のさまざまな見え方を想起する。図16(b)では入力画像が不完全なため、最初三角柱も想起しかけていることが分る。

7. お わ り に

視覚系が、いくつかの機能的モジュールから構成されていることは、ある程度予想されていたことではあるがアーキテクチャとして興味深い。いくつかのモジュールがたがいに情報を交換しあって、全体としてすぐれた機能を実現しているものと思われるが、この点につき、実験と理論の両面から研究することが、今後の大きな課題であらう。

それにしても、ほとんど全く分っていないのが、視覚情報処理における知識の表現と利用である。処理のソフトウェアと言ってもよいかもしれない。若い研究者に突破口を期待したい。

参 考 文 献

- 1) E. A. Deyoe, D. C. Van Essen, "Concurrent Processing Streams in Monkey Visual Cortex", TINS, Vol.11, No.5, pp. 219-226, 1988
- 2) M. S. Livingstone, "Art, Illusion and the Visual System", Scientific American, pp. 68-75, January 1988
- 3) M. S. Livingstone, D. H. Hubel, "Anatomy and Physiology of a Color System in the Primate Visual Cortex", J. Neuroscience, Vol.4, pp. 309-356, 1984
- 4) D. Marr, "Vision", Freeman, San Francisco, 1982
- 5) T. Poggio, V. Torre, C. Koch, "Computational Vision and Regularization Theory", Nature, Vol.317, pp. 314-319, 1985
- 6) 加藤, 大西, 杉江, "バイオリジカルモーションに対する運動立体視機構のモデル", 信学技報, MBE 87-144, 1988
- 7) 安田, 安藤, 大西, 杉江, "物理的に存在しない輪郭線の抽出", 信学論文誌, Vol. J 73D-II, No.6, pp. 906-913, 1990
- 8) 黄, 大西, 杉江, "生体に示唆を得た音源定位システム反響のある環境での単一音源定位", 信学論文誌, Vol. J 71-A, No.10, pp. 1780-1789, 1988
- 9) R. J. Watt, "Scanning from Coarse to Fine Spatial Scales in the Human Visual System after the Onset of a Stimulus", J. Opt. Soc. Am. A, Vol.4, No. 10, pp. 2006-2021, 1987
- 10) 荒木, 大西, 杉江, "大重解像度 F^2G オペレータによるテクスチャの領域分解", 画像理解の高度化と高速化シンポジウム, pp. 175-180, 1989
- 11) A. C. Bovik, M. Clark, W. S. Geisler, "Multichannel Texture Analysis Using Localized Spatial Filters", IEEE Trans. PAMI, Vol.12, No.1, pp. 55-73, 1990
- 12) H. H. Bülthoff, H. A. Mallot, "Integration of Stereo, Shading and Texture", pp. 119-146, in "AI and the Eye", A. Blake, T. Troscianko, eds., John Wiley & Sons, 1990
- 13) T. Poggio, E. B. Gamble, J. J. Little, "Parallel Integration of Vision Modules", Science, Vol.242, pp. 436-440, 1988
- 14) J. Y. Aloimonos, "Unifying Shading and Texture through an Active Observer", Proc. Roy. Soc. Lond. Vol. B 238, pp. 25-37, 1989
- 15) J. Hutchinson, C. Koch, J. Luo, C. Mead, "Computing Motion Using Analog and Binary Resistive Networks", IEEE Computer, pp. 52-63, 1988
- 16) S. R. Lehky, T. J. Sejnowski, "Neural Network Model of Visual Cortex for Determining Surface Curvature from Images of Shaded Surfaces", Proc. Roy. Soc. Lond., Vol. B 240, pp. 251-278, 1990
- 17) G. W. Cottrell, P. Munro, D. Zipser, "Image Compression by Back Propagation: an Example of Extensional Programming", ICS Report 8702, 1987
- 18) N. Sugie, H. Honda, N. Ohnishi, "A Connectionist Model of Associative Storage and Recall of 3D Objects", Proc. IJCNN, Vol.1, pp. 809-814, June 1990
- 19) M. Seibert, A. M. Waxman, "Learning Aspect Graph Representations from View Sequences", pp. 258-264, in "Advances in Neural Information Processing System 2", D. S. Touretzky, ed., Morgan Kaufman, 1990

杉江 昇 (Noboru SUGIE)

昭和 32 名大・工・電気卒。同年通産省電子技術総合研究所入所。昭和 45 バイオニクス研究室長、昭和 53 視覚情報研究室長。昭和 54 名古屋大学工学部情報工学科教授。平成 2 名古屋大学大型計算機センター長。バイオニクス、コンピュータビジョン、医用工学などの研究・教育に従事。工学博士、情報処理学会、エム・イー学会、IEEE 等会員。

(日本ロボット学会正会員)