

TR-A-0047

視覚・認知機構研究における
並列処理計算機 NCUBEの利用について

三宅 誠 乾 敏郎 川人 光男 曾根原 登
Sei Miyake Toshio Inui Mitsuo Kawato Noboru Sonehara

本郷 節之 池田 孝利 中根 一成
Sadayuki Hongo Takatoshi Ikeda Kazunari Nakane

1989. 3. 31

A T R 視聴覚機構研究所

© (株) A T R 視聴覚機構研究所

目次

1. はじめに	1
2. 視覚・認知機構研究と並列処理計算機	2
2.1 ニューラルネット	2
2.2 視覚・認知機構の計算論的研究	3
2.3 ニューラルネットの具体化研究ーニューロコンピューター	4
2.4 並列処理計算機の利用ー画像処理とニューラルネットー	5
3. 並列処理計算機NCUBE	8
3.1 導入の経緯	8
3.2 システムの概略	11
4. 光線追跡法による画像生成	15
5. Sobel-Filtering による輪郭抽出	18
6. ヒトの網膜における情報処理とそのモデル	23
7. エネルギー最小化による画像復元とエネルギー学習	34
8. ネオコグニトロンによる手書き数字認識	47
9. 3層逆伝播学習ネットワークによる画像データ圧縮	53
10. 前注意過程における図形の特徴点抽出モデルと画像復元	63
11. エネルギー学習をする局所並列確率アルゴリズムを用いた輪郭抽出	74
12. おわりに	88

1. はじめに

ATR視聴覚機構研究所では1987年4月15日にハイパーキューブ型並列処理計算機NCUBEを導入、以後視聴覚機構の計算論的研究に使用しほぼ2年が過ぎようとしている。本稿では国内で初めて研究ツールとして購入したNCUBEのハードウェアとソフトウェア、及びATRでの利用実績についてまとめた。

本稿が今後の「並列処理」、「ニューロコンピューティング」、「視聴覚機構」、「ビジョン」、「画像処理・理解」などの研究に新しいアプローチを提供する一助となれば幸いである。

個々の利用実績の記述については、4、5、8章を三宅が、6、10章を乾が、7章を川人が、9章を曾根原が、11章を本郷が、主として担当した。

なお、本稿の各利用例におけるプログラムの大部分は柵CSK（担当：前川雅則氏、大本明氏、米元雅裕氏）の協力を得て作成した。ここに謝意を表す。

また、いつも一緒に研究を進めている、松下博、藤井秀夫、岡本俊明、入江文平、片山正純の研究員諸氏に感謝する。

2. 視覚・認知機構研究と並列処理計算機

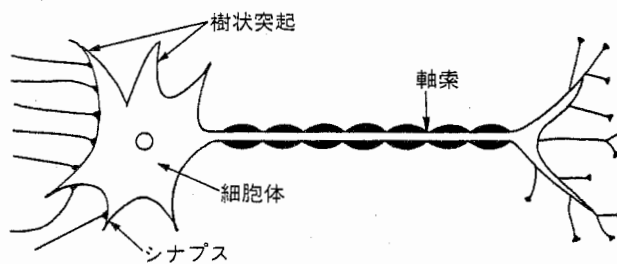
人間の視聴覚情報処理や認知・行動の仕組みを解明しつつ、そこから得られる知見を工学的に応用しようというのが本研究所の主要テーマである。

このような研究においては、実際にニューラルネットの構築原理を導出し、設計図を描くと同時にその応用可能性を示すことが求められる。

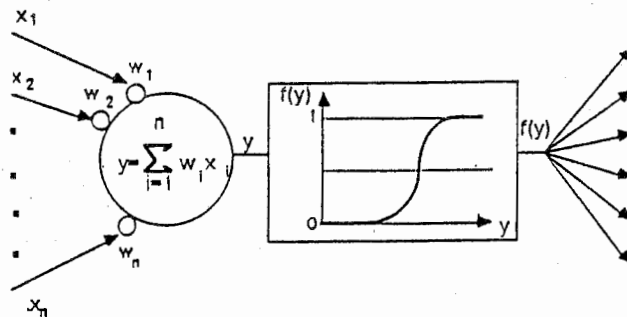
以下、本研究所における研究の内容と並列処理計算機との関わりについて述べる。

2.1 ニューラルネット

人間の視覚情報処理機構ことに脳の構造と機能については、おもに生理学者達によって着実に研究が続けられその基本構造が徐々に明らかにされつつある。神経細胞の数は100億個以上と言われそれらが複雑に結合している。それぞれの細胞(図1)は多数の入力端末とひとつの出力端末を持ち、端末の種類によって信号を送る相手の細胞の出力を増大させたり(興奮性)減少させたり(抑制性)する。細胞は、通常パルス出力を出すのであるがパルス密度によってアナログ変調がかかっており、単純化した信号処理ユニットとしてモデル化すると図1bに示すような非線形処理を行なっているものと推定される。



(a) 神経細胞



(b) 神経細胞モデル

図1. 神経細胞と神経細胞モデル

すなわちごくおおまかに言って、脳は非常に多くのアナログプロセッサが相互に結合した計算機と見なせるが、従来の計算機とは異なり「自己組織能力」⁽¹⁾を持つ。つまり、環境に適応して自己の構造を変えていく柔軟性を有する。我々が成長とか学習と呼んでいる機能はこの自己組織能力に負うところが大きい。

視覚情報処理という視点で考えれば、眼球を介して網膜上に投影された画像は並列的に配置された視細胞でとらえられ電気信号に変換されて脳に送られる。脳では入ってきた情報をさらに分析・統合し、学習、認識、連想といった高度の機能を実現している。例えば、我々は、風景やテレビジョン画像の中から自分の知っている物体や人の顔を、かたち（パターン）として瞬時に認識する。たとえそれらが多少変形したり、ノイズで汚されていても良く似た形のものであれば同じものと解釈出来る。またパターンの全体を、全て同じ重みで受容するのではなく刺激に積極的に働きかけ、興味のある部分や必要とする部分に注目して能動的に情報を処理する。

記憶については、入って来た情報をただ順番に何の脈絡もなくため込むのではなく、良く似たものや関連のあるものをまとめながら、時にはパターンそのものとして時には言葉で、想起しやすい方法で蓄積する。すなわち、従来の計算機原理にはないパターン認識、効率的な記憶方法、あいまいさを許容する柔軟性、能動的情報処理など、種々の機能を有しているのである。

このような優れた生体機能がどのようなメカニズムで実現されているかを探る研究は、すでにニューラルネット（神経回路モデル）の構築という形で進められ着実に成果をあげつつある。ニューラルネット研究は、神経細胞というプロセッサを多数搭載しそれらが相互に結合したシステムがどのように働けば高度の機能を発揮し得るか、という並列階層処理向きの計算原理を探る研究であると言ってもよい。

2.2 視覚・認知機構の計算論的研究

従来のニューラルネット研究においては生理学的知見を参考にして、計算手順やそのモデル化が試みられ多くの成果が上がっている。しかし画像や音声・言語といった実在の対象に関する処理については「何が計算の目標か?」、「その方略は?」といった計算論的視点が不足していたと言っても過言ではない。

人間の視覚情報処理の解明を課題にした計算論的研究の枠組みを最初に与えたのは、1980年35歳の若さで世を去った David Marr である。1982年に刊行された "Vision"⁽²⁾ で彼は、網膜像の強度変化（濃淡値）から実世界の3次元構造を推測することが視覚情報処理系の目標であるにとらえている。従って、視覚の並列・階層的情報処理機構はこの「3次元構造の推測」という目標を達成すべく合理的に設計されているはずである。そこで階層処理の格段では何がどのような目的で計算されているかを明らかにすることが重要である。これは視覚・認知の計算理論とも呼ぶべきもので系の入出力関係の記述と言ってもよい。

ここで重要な点は計算が何らかの暗黙の仮定（拘束条件）のもとでなされているということである。網膜像から3次元構造を推測すること自身が、3次元物体から2次元濃淡画像への写像の逆問題に相当するために、拘束条件なしに計算すれば解が一意に定まらないことが多い。その反対にコンピュータグラフィクスにおける画

像生成の問題は、3次元構造から2次元濃淡画像への写像に関する順方向の問題であり、拘束条件なしに解が得られる。

そこで、この拘束条件についての知識を適切に導入することが求められる。例えば、立体視における左右の網膜像の対応決定過程においては、対象物の奥行きが連続的に変化しているという「連続性」が仮定される。運動の奥行き効果 (kinetic depth effect) のように動きから対象物の構造を復元する過程では、その対象物が剛体 (形が変わらない) という「剛体性」の仮定が必要とされる。後述するマルコフ確率場の考えによる画像復元問題については、各画素の持つ濃淡値がその近傍の画素の濃淡値にのみ依存して定まるという「局所性」を仮定する。

計算理論が明らかにされれば、計算が具体的にどのような手順で実行されているのか、入出力はどのような表現形式が採用されているのかを明らかにせねばならない。例えば、立体視の対応決定過程においては、入力はゼロ交差であり出力は両眼視差であると Marr は考えている。

最後に、得られた計算手順がどのような神経回路で実現 (implementation) されているかが問題になる。従来のニューラルネット研究においても生理学的知見を基礎にしたモデルが提出されている。例えば、後述するネオコグニトロンについては、視覚一次野に関する回路構造的知識をヒントにした「階層的位相ずれ吸収」の機構が備わっている。このような構造についての知識も、計算理論の観点から言えばある種の拘束条件を与えていると解釈できる。つまり、上述したような入力情報そのものに由来する拘束条件と同時に、観察者である人間が進化の過程で獲得して構造化された視覚系そのものからも拘束条件についての知識が得られるのである。

従って、「計算理論」、「計算手順」、「回路構成」の3段階の過程を順を追ってクリアしていくことによるのみ視覚系が理解できるという訳ではなくそれぞれのレベルでの知識を統合したモデルを作るといった新しい研究の枠組みを設定することができる。本研究所においても、このような枠組みの中で、「知識 (拘束条件) を得るための実験や計算理論構築」、「システムとして具体化するための並列処理を機軸とするモデルとその実証」の両側面から研究を進めている。

2.3 ニューラルネットの具体化研究—ニューロコンピューター—

ニューラルネットの特に応用的側面からの研究開発の最も大きな原動力は「並列処理メカニズム」の魅力である。これまでの工学的情報処理方式がノイマン型計算機のアーキテクチャの影響を受け逐次処理に偏っていたため、まだ未開拓の並列処理計算原理がクローズアップされていると言ってもよい。すでにニューラルネットのような並列処理向き計算機構を既存の技術でハード化し、ニューロコンピューターを開発しようという動きも活発化している。

デバイス技術においては、多数のプロセッサを組み合わせることでチップ化することが夢物語ではなくなりつつあり、その技術を生かす計算方式についてはニューラルネットに期待出来るとする認識が高まりつつある。例えば、AT&Tでは、物性物理出身の研究者や生理学研究者が共同し、神経細胞を模した多数のプロセッサを結合してチップ化する研究を始動させた⁽³⁾。カリフォルニア工科大学⁽⁴⁾やMIT⁽⁵⁾

でも同種のL S I化研究が始められている。Hecht-Nielsenによれば⁽⁶⁾、ニューラルネットとして新しい機能を実現できる。基本原理は着実に蓄積されつつあるが、いずれも実用化レベルのシステムにはなっていない。実用化に当たっては、システムとして組み上げる高密度並列ハードウェア開発が不可欠であり、3つのアプローチが考えられるという。示唆される点が多くその概要を紹介する。

①光演算器の利用 ②V L S I化 ③仮想ハード化

①は、レンズの組み合わせやホログラフイによって相関等の演算を行なわせるという方式である⁽⁷⁾。このアプローチは、アナログ性とリアルタイム性を原理的に確保出来る。しかし、細胞や結合という単位で構築されるニューラルネット計算原理を生かせるかどうかは今後の課題である。

②は、A T & T⁽³⁾等で試みようとしているアプローチで、近い将来の実現性が高い⁽⁸⁾。しかし現状ではモデルの多くが開発途上にあり、逐一チップ化する労力が問題となる。

③は、とりあえずニューラルネットがリアルタイムに近い形で動くデジタルコンピュータを用意することが得策、とするアプローチである。計算機を並列処理型に置き換え、より実用レベルに近づいた実験で応用可能性を見いだそうという方略である。

Hecht-Nielsenは、③のような方向で利用出来るアクセラレータとしてマークⅢ、マークⅣを相次いで発表した後、ニューロコンピュータ社という会社を作り、1ボードアクセラレータとそのソフトを製品化している。その他シナプティクス、ニューラルテックといったベンチャーも誕生している⁽⁹⁾。国内においてもこの種の開発が盛んになってきている⁽¹⁰⁾。本稿で紹介する並列処理計算機N C U B Eを利用した研究もこの仮想ハード的アプローチの枠組みにあると言えよう。

2.4 並列処理計算機の利用—画像処理とニューラルネット—

並列処理向き計算原理を持つのは、ニューラルネット、画像処理、流体力学、計算物理、地質学など数多い⁽¹¹⁾。それらの分野は、実世界を対象にした非常に解きにくい、あるいは解けても時間のかかる問題を抱え計算機による視覚化(visualization)によって実体に挑もうとする特徴を持っている。

本節では、人工知能研究に新風を吹き込む可能性の強い画像処理やニューラルネットを中心に並列処理計算機との関係について述べることにする。

2.4.1. 画像処理

画像についてはこれまでもパイプライン処理やベクトル化といった並列化が試みられているが、データそのものの並列性を生かした処理が行なわれているとは言い難い。逆に、瞬間的にいっせいに入ってくるデータの並列性と膨大さという性質が大きなネックになっていた。すなわち画像を共用メモリーに行列としてストアし、

1画素ごとに取出してCPUが演算し、その結果をまた別の行列に置くといった方式を取らざるを得ない。ソフトウェアでは、データを2重のDOLoopで処理することになりその計算時間は膨大である。また、データの並列性をあえていったん放棄し、逐次処理的に扱うという翻訳の手間を求められていた。

メモリー容量も問題となる。テレビジョン画像でも数十万画素を一度に扱うことになり、1画素を1ワードでデータ表現しようとする、数百キロワードを1画像行列で占有してしまう。このようなアーキテクチャは、メモリー価格がプロセッサのそれに比して非常に安かったという理由で生まれたものであるが、高速性を要求される新たな応用分野が拓けてこない。それでも、コンピュータグラフィクスなど画像生成については方式研究が進み、種々の並列処理向き原理が考案されている。その一つ光線追跡法は、1画素ごとに視点に到達する光を物体との交点における光学的性質（反射、屈折、透過）に従って逆方向に探索し、現実感のある画像を生成する方法である。画素ごとに独立した操作の繰り返しであるため、並列処理が得策とされる⁽¹²⁾。光線追跡法についてはすでにコネクションマシン、NCUBEで画像生成が行なわれている^{(13), (14)}。

2.4.2. ニューラルネット

一昨年のスノーボード会議の後、ビジネスウィーク誌は初めてニューラルネット特集を組み、コネクションマシン、iPSC、NCUBE、Tシリーズ等の並列計算機を新しい研究の道具として紹介した⁽¹⁵⁾。

従来は、応用を意図した規模の大きいシミュレーションを行なうには処理速度などに限界があり下記のような要請に対する解決策が求められていた。

- (1) 音声や画像などの実データを扱えるモデルを構築してその能力を評価すること
- (2) 計算が時間的に高コストとなる大規模モデルを実行する手だてを講ずること

このような要請の下に、並列計算機が新たな研究の道具として利用され始めた。ニューラルネットの特徴は『相互に結合した多くの処理ユニットが同時並列的に働いて目的とする結果を得る』ことにあり、この点でニューラルネットと並列計算機は非常に重要な共通の性格を有すると考えられるのである。

2.5 文献

- (1) 福島邦彦: 神経回路と自己組織化, 共立出版(1979)
- (2) D. Marr: Vision, Freeman (乾, 安藤訳: ビジョン-視覚の計算理論と脳内表現, 産業図書(1987))
- (3) L. D. Jackel, H. P. Graf, R. E. Howard: Electronic Neural Network Chips, Applied Optics, 26, 23, pp. 5077-5080(1987)
- (4) M. A. Sivilotti, M. R. Emerling, C. A. Mead: VLSI Architectures for Implementation of Neural Networks, AIP Conf. Proc. 151, pp. 408-413(1986)
- (5) J. P. Sage, K. Thompson, R. S. Withers: An Artificial Neural Network Integrated Circuit Based on MNOS/CCD Principles, AIP Conf. Proc. 151, pp. 381-385(1986)

- (6) R.Hecht-Nielsen: Performance Limits of Optical, Electro-Optical and Electronic Neurocomputers, TRW Rancho Carmel AI Center Report, pp.1-45 (1986)
- (7) N.H.Farhat: Optoelectronic Analogs of Self-programming Neural Nets: Architecture & Methodologies for Implementing Fast Stochastic Learning by Simulated Annealing, Applied Optics, 26, No.23, pp.5093-5103(1987)
- (8) 土屋泰一他: 鳴動するニューロチップ, 日経マイクロデバイス, 1988.7月号, pp.43-89(1988)
- (9) E.Rosenfeld: INTELLIGENCE, April, pp.1-3(1987)
- (10) 日経エレクトロニクス, 1988.6.27号, No.450, p.92(1988)
- (11) G.C.Fox, S.W.Otto: Algorithms for Concurrent Processors, Physics Today, 37, 5, pp.50-59(1984)
- (12) 安田孝美, 横井茂樹, 鳥脇純一郎, 鶴岡信治, 三宅康二: 透明物体表示のための改良光線追跡法, 情報処理学会論文誌, 25, 6, pp.953-959(1984)
- (13) D.Hillis: The Connection Machine, Scientific American, 256, 7, pp.108-115(1987)
- (14) 三宅誠, 安田孝美, 鳥脇純一郎: 並列処理計算機による画像の高速生成, 昭和62年電気関係学会関西支部連合大会論文集, S64(1987)
- (15) O.Port: Computers that Come Awfully Close to Thinking, Business Week, June 2, pp.66-71(1986)

3. 並列処理計算機 NCUBE

1986年6月のビジネスウィーク誌がニューラルネット研究ツールとしてハイパーキューブ型並列処理計算機を紹介して以来、それらの処理性能、応用などについての議論が高まった。

ATRにおいても、並列処理向きモデルを効率的にシミュレーションし、時にはリアルタイム処理や会話的処理によってその工学的応用を示せる計算機の導入を計画しつつあった。以下、機種選定を含む導入および整備の経緯について述べる。

3.1 導入の経緯

選定に入った1986年秋には表1に示すような商用機が登場しつつあった。本表には、機種、メーカー、最大搭載プロセッサ数だけを示しているが、その他の仕様については日経エレクトロニクス誌(1986.11.3号)の記事に詳しい⁽¹⁾。

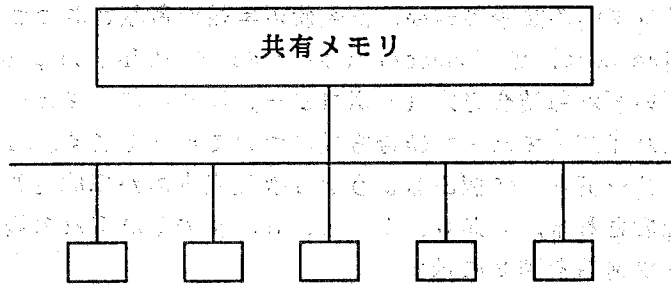
	型名	開発元	プロセッサ数
密結合	FX/8	Alliant	8
	Balance	Sequent	30
	Multimax	Encore	20
	FLEX/32	Flexible	40
疎結合	Connection Machine	Thinking Machine	65536
	NCUBE	NCUBE	1024
	T-Series	Floating Point Systems	16348
	iPSC	Intel	128
	Butterfly	BBN	256

表1. 商用並列処理計算機(1987.1月)

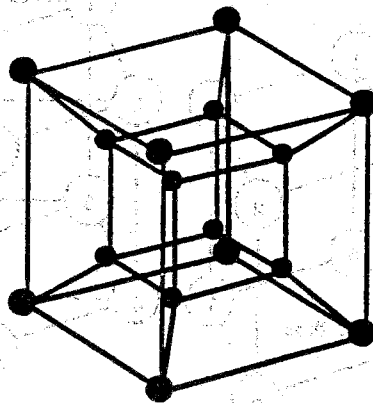
密結合型とは、複数のプロセッサで一つのメモリを共有するタイプ(図1a)で、OSは共有メモリ上に置き各ローカルプロセッサで共用する。共通バスのアクセス頻度が非常に高くなることから、各ローカルプロセッサにはキャッシュメモリを付

加する。

これに対する疎結合型とは、複数のプロセッサを複数のバスによって互いに接続したタイプ（図1b）であり、その接続形態にはメッシュ、バタフライ、ハイパーキューブなど様々な結合方式がある。密結合型との本質的差異は、メモリを各ローカルプロセッサに分散配置する点である。各ローカルプロセッサにはローカルメモリやI/O機能が付加され、OSも全てのローカルメモリに置かれる。



(a)密結合型



(b)疎結合型

図1. 並列処理計算機のプロセッサ結合方式

密結合型については1985年初頭から商用化が始まり、稼働実績も上がりつつあった。例えばアライアント社のFX/8については、ATR自動翻訳電話研究所で音声認識等の研究に使用すべく1986年秋にはすでに導入を決定していた。

一方疎結合型については、わが国における使用実績もほとんどゼロに等しく研究開発レベルとの認識が先行していた。多数のプロセッサ間通信のロスはどう軽減するのか、軽減できなくともバランスをどう取るのかといった問題についての不安があったのである。

しかし、共有バスの競合によって数十台のプロセッサで理論的限界に到達してしまう密結合型に比べて数百、数千、数万といったプロセッサを非同期で独立に動かすことに多大の期待が寄せられていた。ことにニューラルネット型計算モデルは通

常数千、数万のエレメントが相互作用する構造になっており、データレベルの並列性を利用できる疎結合型の先端的アーキテクチャは大きな魅力であった。従って機種種の選定においては、「将来性、先端性の疎結合」か「実績、安定性の密結合」かという選択を迫られた。

息の長い基礎研究としての広い枠組みで「新しいタイプの並列計算原理を新しいタイプのアーキテクチャで実証する」という課題に挑戦しようとするならば、疎結合型について海外も含めたユーザーの利用状況を確認しておくことが先決であると考えた。ヒアリング等の調査の結果、研究使用実績が蓄積されつつある疎結合型は Thinking Machine Corp. の Connection Machine と NCUBE Corp. の NCUBE の 2 機種に絞られた。いずれも結合方式（トポロジー）はハイパーキューブである。

ここで簡単にハイパーキューブ結合方式について述べておく。n次元のハイパーキューブでは、ノードが 2^n 個あるような n次元立方体が構成され、その各頂点にプロセッサが配置される。nが0、1、2、3、4のそれぞれの場合についてそのハイパーキューブ構造を図2に示す。

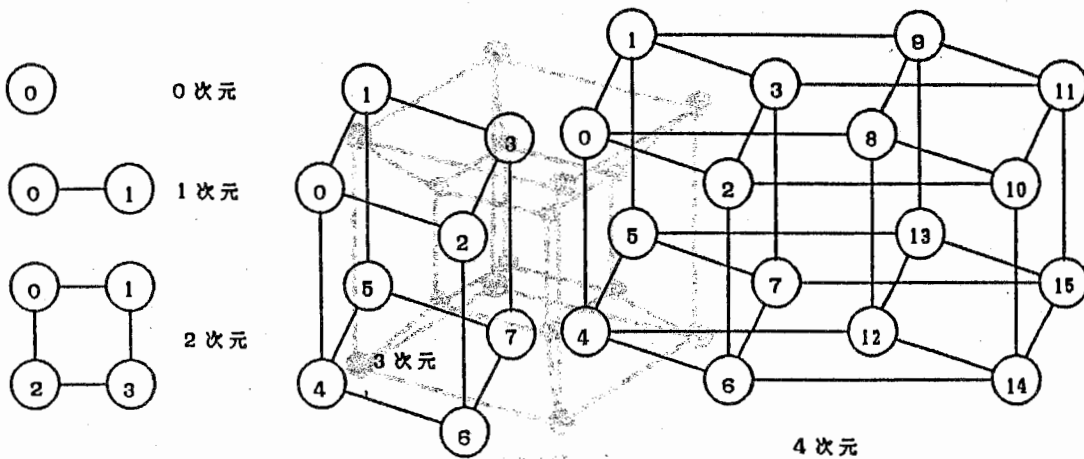


図2. ハイパーキューブ結合方式

図2から推察されるようにハイパーキューブ構造は、各ノードプロセッサから他の全てのプロセッサと等距離で通信できるような理想的構造からは程遠い。つまり、手が出ている隣の頂点（3次元ハイパーキューブの場合は隣の3個）とは短距離通信が可能であるが、それ以外の頂点のプロセッサとは別の頂点を経由して通信せねばならず通信負荷が大きい。これがハイパーキューブのネックと言われている。しかし、相手のプロセッサによって通信距離の差が大きくなってしまっても、その差はn次元座標上での差にとどまり、例えばメッシュの左下と右上間の距離と隣接点の距離の差ほど大きくはならない。

さて Connection Machine は疎結合型の中でも最も粒度の高い (fine grained) 計算機であり、65536個という「数」と引き換えに各々1ビットの簡単な計算処理を

行なうプロセッサで構成される。従ってノードプロセッサの負荷を小さくすることによってはじめてその能力を最大限に発揮できる。もしこのアーキテクチャに非常に良くマッチする計算原理を開拓できれば画像や音声の実時間処理も夢ではない。実際 MIT AI ラボ等では両眼で取り込んだ画像情報を手がかりにした実時間運動制御システムを構成している⁽²⁾。以上のようなマシンの性格から次ぎの2点に留意して使用することが求められる。

- (1) 1ビット、65536個を生かすに適した問題の選択と並列計算原理の構築
- (2) 各プロセッサの負荷と通信時間のバランスを取る工夫

もちろんこの2点は疎結合型に共通の問題であるが NCUBEのように数百個の32ビットプロセッサで構成される粒度の低い(course grained)マシンでは、かなり複雑な処理を各プロセッサに担当させることができると同時にプロセッサ負荷に比べて通信時間が小さく、適合する問題を選択する範囲が広がるとの感触を得た。NCUBEの具体的ユーザーが存在する分野としては、石油探査、流体、気象などにおけるグラフィックシミュレーションなどが挙げられ、それらの成果に関する報告を入手した⁽³⁾。

また NCUBE の原型 Cosmic Cubeを開発したカリフォルニア工科大学の G.C.Fox 教授および NCUBE Corp. の S.R.Colley 社長と面談し、NCUBE型の低粒度マシンが Connection Machine型の高粒度マシンをターゲットとしている訳ではなく、低粒度マシンとしての長所を生かした設計指針を持っていることを確かめた。さらに実時間性を狙う次世代バージョンは4倍から8倍のプロセッサ数を見込むとの考え方が提示された。

上述の事前評価に基づき、主として下記2点の理由から NCUBEを選択した。

- (1) 国内初の疎結合型の導入にあたって先端性と安定性のバランスがとれている
- (2) Connection Machineについては、Thinking Machine社が日本への輸出を考えていない時点で導入スケジュールの青写真が描きにくかった

3.2 システムの概略*

NCUBE/10 は、公称 2.5MIPS, 0.4MFLOPS の能力を持つ32ビットプロセッサを最大1024個実装できる完全分散処理型の計算機である。NCUBEは密結合型計算機と異なり、大規模なパイプライン/ベクタ処理による高速化も、プロセッサ間での同期を取りながらの並列処理も行なわない。NCUBEの各ノードは、ユーザーのプログラムからのみ他のノードやホストプロセッサとの通信を行なう。従って通信のプロトコルもユーザーにまかされた形態になっている。図3にシステム外観を、図4にシステムの仕様概略を示す⁽⁴⁾。

3.2.1 ホストプロセッサ

NCUBE/10のホストプロセッサはインテル社製 80286 で、ノードプロセッサとユーザーとのインターフェースの役割を担っている。ホストプロセッサは、まず必要

* 本節(3.2)の記述の主要部分は当所視覚研究室城和貴研究員によるものである

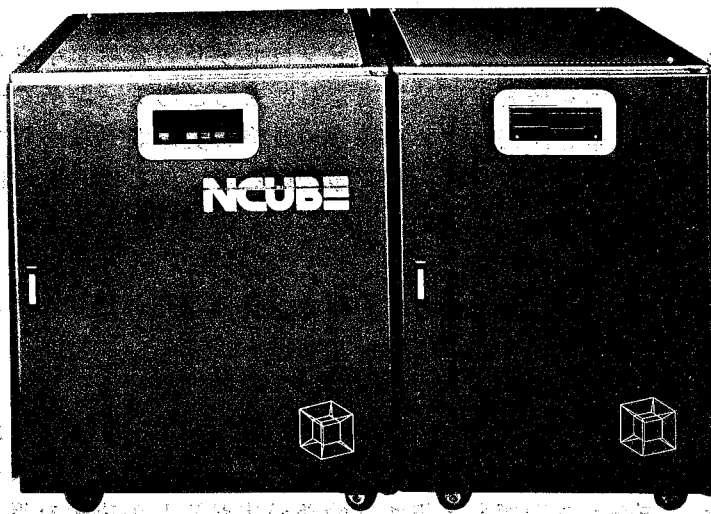


図3. システムの外観

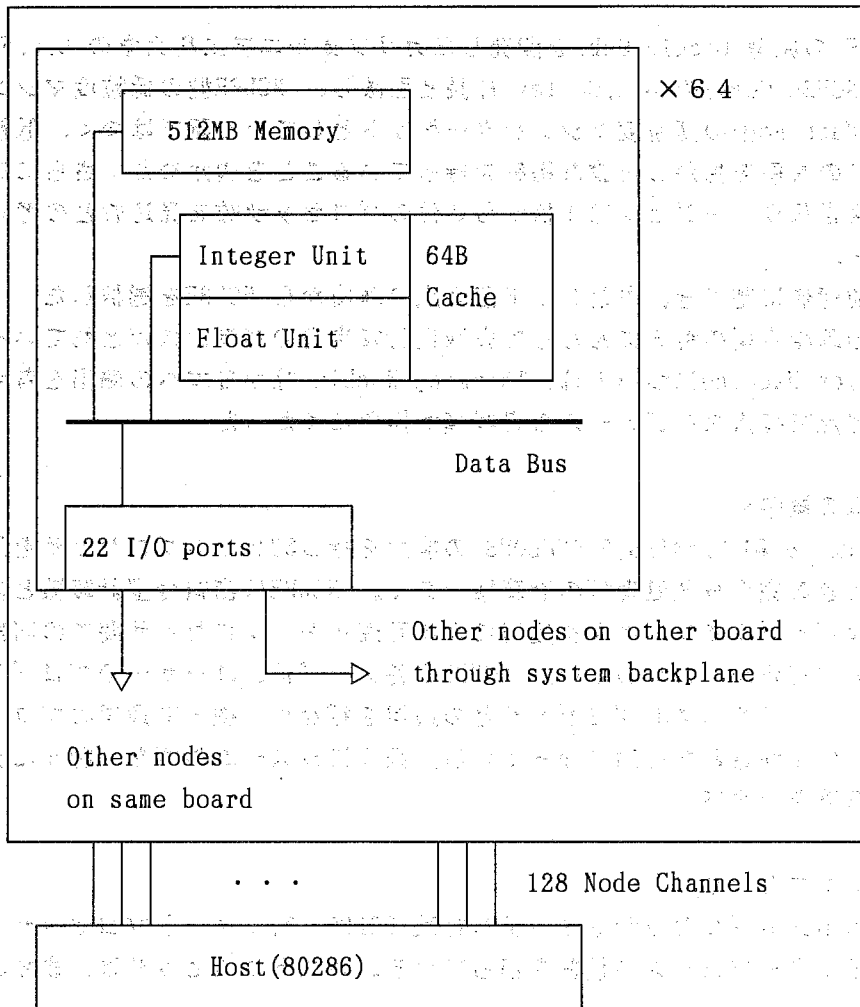


図4. NCUBE/10のシステム仕様概略

な数のノードプロセッサをアロケートし（ただしアロケートするノード数は $2^n \leq 1024$ に限定される）、ノードプロセッサ用プログラムをロードし、必要に応じてデータを転送する。各ノードプロセッサはホスト経由でディスクやユーザーからのデータを受け取る以外は全く独立に機能する。なお、ホストプロセッサは Axis と呼ばれる Unix に似た OS で管理されている。また、ノードとの通信は Vortex と呼ばれる Axis の拡張システムで管理される。

3.2.2 ノードプロセッサ

NCUBE/10のノードプロセッサは、8MHz の 32ビットマイクロプロセッサに高速の I/O チャンネルを付加した形態になっている。従って、各ノードプロセッサは通常の数値計算、データ転送、制御命令の他にノード間通信を行なうことができる。

各ノードプロセッサは 512KB のローカル物理メモリを持ち、物理メモリシステムで動作する。また、通信用の Vertex と呼ばれる 32KB のシステム以外は OS を持たず、ユーザープログラムがホストからダウンロードされると直ちにそのプログラムを実行する。

レジスタについては、汎用レジスタが 16 個、プロセッサレジスタが 12 個、I/O レジスタを 22 個持ち、すべて 32 ビット長である。汎用レジスタは整数計算、小数計算の他にアドレスレジスタとしても使用される。倍精度計算を行なう時は 2 個の汎用レジスタを一組にして使う。プロセッサレジスタは、スタックポインタ、プロセッサ ID、フォルトレジスタをはじめとする各種プロセッサ状態の貯蔵や I/O 割り込みに使用される（ただしここで言う I/O とは、プロセッサ間通信を指す）。I/O レジスタは、他のノードプロセッサとの通信を行なう際のバッファアドレスとメッセージ長を示すのに用いられる。

3.2.3 その他

NCUBE は前述のように大規模なパイプラインベクトル化は行なっていないが、各ノードプロセッサにおいて以下のような高速化の工夫がなされている。

- (1) 小数計算時に次のオペランドのプリフェッチを行なう疑似パイプライン
- (2) 32 バイトのインストラクションキャッシュ

ノード間通信は 22 個の DMA・I/O チャンネルによって行なわれる。チャンネル 0 から 9 までは他ノードからの入力に、32 から 41 は出力に、31 と 63 はホストおよび I/O ポート（例えばグラフィックスのフレームバッファ）との入出力に使用される。チャンネル 0 から 30 と 42 から 62 は将来の拡張に備えてリザーブされている。各チャンネルは送受信のためのメッセージアドレスとメッセージ長を示すレジスタを 2 個ずつ持っている。各チャンネルの DMA 転送速度は 180MB/S で、プロセッサレジスタの状態に依ってメッセージの送受信、インタラプト、ペンディングウエイト等を、ノードプロセッサと独立して行なう。各ノードはブロードキャスト命令により 11 の出力チャンネルから同時に同じメッセージを DMA 転送できる。

なお ATR においては昭和 62 年 4 月の時点で 256 個のノードプロセッサを実装し、その後昭和 63 年 2 月に新たに 256 個を追加現在 512 個で稼働中である。

3.3 文献

- (1) 稲葉則夫:期待と不安を乗せて飛び立った並列処理コンピュータ,日経エレクトロニクス,407,pp.119-129(1986)
- (2) M.Drumheller,T.Poggio: On-Parallel Stereo, Proc. IEEE Conf. on Robotics(1986)
- (3) D.Hillis:The Connection Machine, MIT Press(1985)
- (4) In: Caltech:Concurrent Supercomputing Initiative at Caltech Information Package, C³P-428(1987)

(4) NCUBE Corp.: NCUBE Handbook 1.0(1987)

4. 光線追跡法による画像生成⁽¹⁾

4.1 はじめに

(ニューロコンピュータ⁽²⁾、流体力学、理論物理、画像処理など並列処理向き問題を抱える様々な分野において、並列処理計算理論の構築とそれらを高速実行できる新しい計算機に対する需要が高まっている。

最近では、互いに通信可能な多数のプロセッサを同時に働かせて高速化する計算機の実用化も徐々に始まっている⁽³⁾。本章では、並列独立に分割処理させるのに適した光線追跡法⁽⁴⁾をハイパーキューブ(超立方体)型並列処理計算機で処理した結果について報告する。

4.2 並列処理計算機

パイプライン処理やベクトル処理も並列処理の1種であるがそれとは別に、多数のプロセッサを搭載した並列処理計算機が商品化され始めた⁽³⁾。今回使用したシステムは、NCUBE と呼ばれるハイパーキューブ型並列処理計算機であり、ホストとは別に、32ビットのプロセッサ 256個を搭載している(プログラム作成当時)⁽⁵⁾。

4.3 光線追跡法による透明物体表示

4.3.1 光線追跡法

スクリーン上の各画素ごとに視点に到達する光を各物体との交点における光学的性質(反射、屈折)に従って逆方向に探索し、現実感のある表示画像を得る方法である。独立した操作の繰り返しであるため、並列処理が得策とされる⁽³⁾。

4.3.2 並列処理

NCUBE システム本体は最大1024個のノードプロセッサを搭載可能であるが、今回は、図1に示すように 256個のプロセッサを持つシステムで分割処理しながら計算結果をグラフィクス装置に表示した。

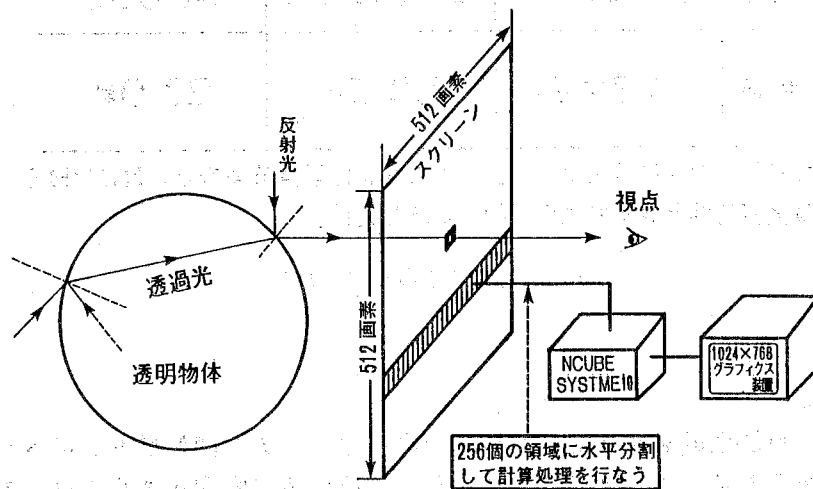


図1. 光線追跡法の並列処理

4.3.3 実験結果

図2のような9個の透明球の生成時間を表1に示す。表示も含めて汎用大型コンピュータを上回るスピードが得られた。また、汎用大型コンピュータでは、いったんディスクに結果を蓄えるため、表示プロセスが計算処理と切り離されてしまう。本システムでは、処理を終えたボードプロセッサは直ちにその結果を表示するので、表示を含めた一連の処理時間を大幅に短縮することが可能となる。

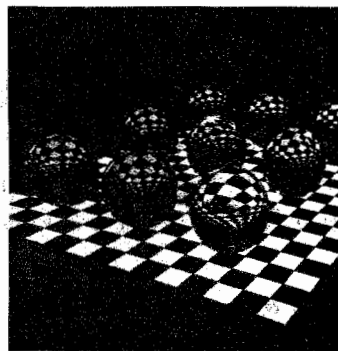


図2. 画像生成結果

計算機名	N社マイクロコンピュータ	F社汎用大型コンピュータ	NCUBE
処理時間	132分	50.5秒	27.0秒

但し、マイコンと大型汎用コンピュータの場合は計算時間のみ。NCUBEの場合は計算時間と表示処理の時間を加算した値。

表1. 処理速度の比較

4.4 おわりに

本稿では、画像生成を並列処理計算機で高速化し、実時間処理やアニメーションへの利用可能性を示した。今後このような計算機の応用範囲をさらに拡大するには、並列計算向き問題を解くための計算理論の構築と蓄積が重要課題である。

4.5 文献

- (1) 〇三宅誠, 安田孝美, 鳥脇純一郎: 並列処理計算機による画像の高速生成, 電気関係学会関西支部連合大会講演論文集, 2A-1-5, pp.137-140(1987)
- 〇三宅誠: ニューラルコンピューテーション, 信学会誌, Vol.70, No.12, pp.1255-1262(1987)
- 〇S.Miyake, T.Inui: Neural Computation Research for Vision, Cybernetics and Systems'88, Kluwer Academic Publishers, pp.1007-1014(1988)
- (2) 三宅誠: 日経マイクロエレクトロニクス, 1987年3月号, p.98
- (3) 稲葉則夫: 日経マイクロエレクトロニクス, No.413, p.159(1987)
- (4) 安田孝美, 横井茂樹, 鳥脇純一郎, 鶴岡信治, 三宅康二: 透明物体表示のための改良光線追跡法, 情報処理学会論文誌, Vol.26, No.4, pp.591-599(1985)
- (5) NCUBE社: NCUBE HANDBOOK Ver.1.1(1986)

5. Sobel Filtering による輪郭抽出^{(1)*}

5.1 はじめに [1][2][3][4][5]、本論文「並列処理による画像認識」(高橋裕子)は、MIMD構成並列処理コンピュータNCUBE⁽²⁾の画像処理への適用性を検討した。本章では画像認識の前処理として輪郭抽出をとり上げ、従来からよく用いられる Sobelの1次微分についてNCUBEの能力を検証する。

5.2 輪郭抽出処理

輪郭やエッジの抽出は神経生理学的にみても認識過程における重要な要素であることが示唆されており^{(3)・(4)}、原画像の豊富な情報を含むことが多い。従来から工学的にも種々の方法が提案されているが⁽⁵⁾、ここでは、並列処理に向き比較的良好な結果を示すと言われる Sobelの1次微分オペレータを採用した。また輪郭処理の前処理としてスムージングによる雑音除去を施した。

5.3 NCUBEと処理方法

NCUBEは1個のホストとハイパーキューブ型に接続された複数の(当時256個)32ビットマイクロプロセッサから成る並列処理計算機である(図1)。

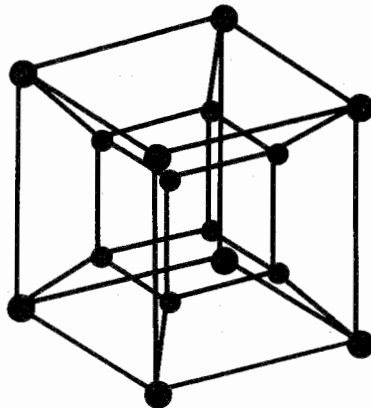


図1. 16(=2⁴)個のノードプロセッサをハイパーキューブ型につないだ例

使用した画像の画素数は 256×256とし、各ノードに1ライン(256画素)分の計算処理を割り当てた。

ホスト側のプログラムは、各々のノードに計算法(独立したプログラム)と必要な画像データ(3ライン分)を送る。各ノードプロセッサは独立に定められた処理を遂行し、終了しだい結果を返す。ノード側ではプログラムがロードされると直ちに処理を開始する。今回の実験では、初めに雑音除去とその結果の受渡し(ノード間通信)を行なう。次いで Sobelの1次微分(並列処理)を施し、処理結果をホストに返す(図2)。

*本実験は高橋裕子(NTT HI研)が S.62.12.1-S.63.1.14 に実習生として担当した。

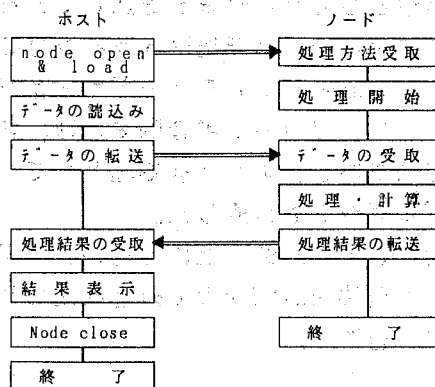


図 2. 処理の手順

以下実際のプログラム作成について概説する。

ホスト側では、パラメータの読み込み、画像データの読み込み、ノードプロセッサのプログラムロード、各ノードに必要なデータ転送、さらにノードからのデータ受け取り及びその結果の表示を行なう。ノードプロセッサはロードされたプログラムと与えられたデータを使って独立にその処理を行なう。今回はノードの処理として Sobel の 1 次微分を行なった。

256×256, 8bit の画像データに対し 256 の各ノードに 1 ライン (256 画素) 分の処理を割り当てた。それぞれのノードには 3 ライン分のデータを送った。3×3 近傍のデータのみが必要な Sobel の 1 次微分の場合、各ノードは完全な並列処理を行なうことができる。従って、このプログラムではノード・ノード間通信は行なわない。

Sobel の 1 次微分を施すにはまず求めたい画素の 3×3 近傍の値に、図 3 に示す Sobel 演算子 d_x , d_y を掛けてそれぞれ加算し、 d_1 , d_2 とする。絶対値の和 $|d_1| + |d_2|$ または $\sqrt{|d_1|^2 + |d_2|^2}$ を求める値とする。今回のプログラムにおいてはこの 2 種類の方式のうち絶対値の和を採用した。

$$d_x = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix} \qquad d_y = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}$$

図 3. Sobel 演算子

プログラムのフローを図 4 に示す。ノードプログラムは無限ループになっているが、ノードにデータが送られるまでその時点で待ち状態となる。ノードプログラム終了 (node close) が発せられた時 (または全ての処理が終了した場合)、ノードのプログラムは終了する。

5.4 結果

図5に原画像と処理結果を示す。また、表1に処理速度とMIPS値概算を示す。それぞれのノードは公称2MIPS, 0.5MFLOPSの性能を持つ。しかしながら単純計算による値512MIPSは得られなかった。これは、ホストによるプログラムロードやホスト・ノード通信が他のノードを介して行なわれることや、公称値ほど実際の能力が発揮できていないなどの理由によるものと考えられる。通信および処理に要する時間は個別に計測し難いため、各ノードの処理の繰り返し回数を増加させながら所用時間を実測した(表1)。その結果、システムの能力は通信に要する時間の減少につれて増大し、約168MIPSに近づくと思われる。Sobelの1次微分のみ処理の場合には通信ロスが無いので速度は向上し、約250MIPSに近づく。



図5. 原画像(256×256,8bit)および Sobelの1次微分の結果例

回数	1	10	30	50	100	500	回
load+データ転送	2	3	3	3	3	3	秒
処理+結果受取	1	2	6	10	21	102	秒
合計	3	5	9	13	24	105	秒
MIPS値概算(*)	9	66	114	132	138	163	MIPS

(*注. MIPS値はVAX8550(公称6MIPS)で同等の処理を行った時の処理時間から逆算)。

表1. ノードにおける処理の繰り返し回数による処理時間の変化とMIPS値概算

5.5 おわりに

画像処理の計算原理によってはNCUBEの性能を有効に活用できることを確認した。本検討が画像認識・理解などの複雑な方式を実証する際の基礎資料となれば幸いである。ところで、今回のプログラミングでは、NCUBEのハイパーキュー

ブ構造利用した並列化を行っていない。今後ハイパーキューブ構造に合わせた並列化手法が開拓されることが望まれる。

5.6 文献

- (1) 高橋裕子,三宅誠,梅田三千雄:並列計算機 (NCUBE) の画像処理性能の評価,昭和63年度電子情報通信学会春季全国大会講演論文集,D-420(1988)
- (2) NCUBE Corp.:NCUBE Handbook, 1.0(1986)
- (3) D. Marr:Vision,W. H. Freeman and Company(1982). 乾,安藤訳:ビジョンー視覚の計算理論と脳内表現ー,産業図書(1987)
- (4) 福島邦彦:視覚の生理とバイオニクス,電子通信学会(1976)
- (5) 田村秀行:コンピュータ画像処理入門,星雲社(1985)

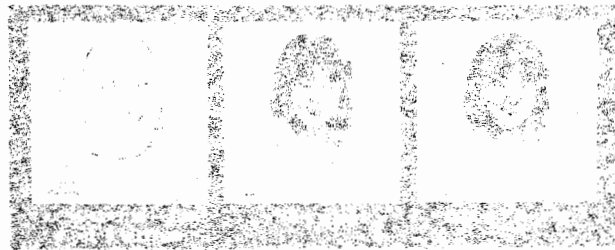


図 5.6 顔の認識のための特徴抽出の過程を示す3つの段階の画像

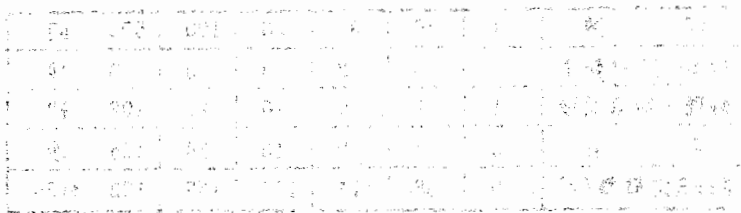


図 5.7 顔の認識のための特徴抽出の過程を示す3つの段階の画像

図 5.8 顔の認識のための特徴抽出の過程を示す3つの段階の画像

図 5.9 顔の認識のための特徴抽出の過程を示す3つの段階の画像

6. ヒトの網膜における情報処理とそのモデル

ヒトの網膜における情報処理様式を完全に理解するためには、個々のニューロンの時空間特性を調べるだけでなく、ニューロンの受容野密度をも調べる必要がある。なぜなら、網膜を一つの情報処理システムと見たとき、ニューロンの時空間特性は画像フィルタ特性に対応し、ニューロンの受容野密度は情報の標本化頻度に対応すると考えられるからである。本研究の目的は、心理物理学の実験データに基づくヒトの網膜モデルを提案し、それによってヒトの網膜のX型、Y型、W型ニューロンの受容野密度を推定することにある。

6.1 心理物理学的知見

被験者にスポット光を提示し、検出閾（光覚閾）を測定すると、スポット光の面積の増加とともに閾値が低下することが知られている。通常、検査光の面積と検出閾の関係は両対数グラフで示される。このようなグラフを閾値-面積曲線(area-threshold curve)と呼ぶ。検査光が小さいときの閾値は面積の増加とともに（両対数グラフでは）-1の傾きで低下する。これは閾値と刺激光の面積が逆比例していることを意味する。すなわち（閾値）×（検査光の面積）=一定、という関係が成立しているのである。これは何らかの機構によって検査光のエネルギーが完全に加算され閾値が決定されていることを意味しており、空間加重が完全であると言う（complete spatial summation）。完全な空間加重が成立する範囲は単に空間加重領域（spatial area）とも言われる。検査光をさらに大きくすれば傾きは-1より小さくなる。しかし傾きは0ではなく部分的に空間加重が成立している（partial summation）。さらに検査光の面積を大きくすると閾値は検査光面積によらず一定になる（non-summation）。Inuiら^{(1),(2)}は、空間加重領域の大きさが、網膜偏心率とともにどのように変化するかをしらべた。背景輝度は10asbであり、検査光は200ms間パルス状に提示された。その結果、空間加重領域の直径d(min)は、網膜偏心率E(deg)と直線関係にあり、

$$d=0.76E+4.2$$

であった。すなわち、中心窩では直径約4'、網膜偏心率が10°の位置では直径12'の範囲で空間加重が成立した。Okamotoら⁽³⁾は、検査光の輝度の対数が時間とともに線形に増加するようにして同様の実験を行なった。この提示条件では時間的に急激な変化が見られない。その結果、空間加重領域の直径d(min)は、やはり網膜偏心率E(deg)と直線関係にあったが、その勾配は小さく

$$d=0.2E+2.2$$

であった。前述のパルス状に提示した条件、およびゆっくりと輝度を上げる条件で得られた空間加重領域の直径はそれぞれ、サル⁽⁴⁾の網膜神経節細胞のY型、X型の受

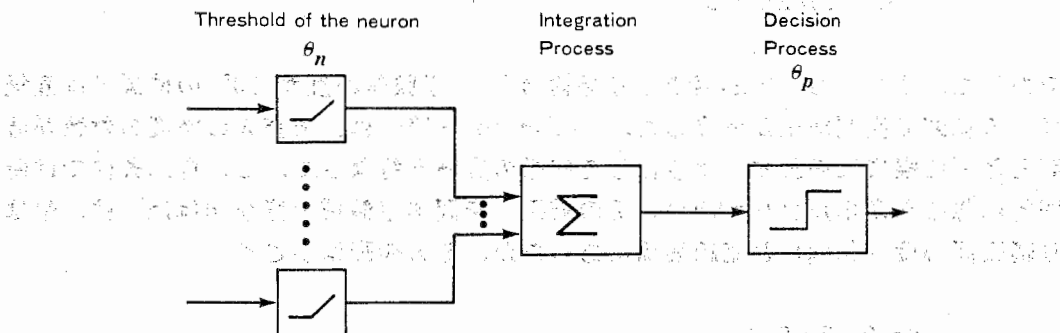
容野中心部の直径にほぼ等しい大きさであった。乾ら⁽²⁾は、抑制効果が働く範囲、すなわち抑制野(inhibitory field)の大きさと網膜偏心率の関係をしらべた。その結果、抑制野の直径は網膜偏心率とともに増大することが明らかにされた。すなわち、中心窩では直径 12°、網膜偏心率 9° の位置では直径 40° であった。この結果を、前述の Y 型の空間加重領域の大きさと比較すると、網膜偏心率によらず、抑制野の直径は空間加重領域の直径の約 3 倍であることが明らかにされた。以上の事実から、検査光が提示されたごく近傍では光エネルギーが加算され、その周辺部では逆に検査光に対し抑制効果が生ずることがわかる。

6.2 問題の設定

まず初めに、上記の閾値-面積曲線において不完全加重領域をどう捉えればよいのが問題である。閾値に関してはいくつかのモデルが考えられるが、不完全加重領域のデータをうまく説明できるものはこれまでなかった。したがって各々の網膜部位における閾値-面積曲線をシミュレートできる網膜モデルを考案することが第 1 の目的である。次に、網膜モデルによって、個々の受容野特性のみならず、受容野密度をも推定し、網膜情報処理の不均一性を定量化することが第 2 の目的である。

6.3 モデル“TAMIT”の構成

TAMIT(Total Activity Model for Increment Threshold)は大きく 2 つの処理部分から構成される。第一の処理は網膜神経節細胞(RGC)の情報処理に対応するもので、第二の処理はすべての RGC の出力の加算を行なう部分である。TAMIT の構成を図 1 に示す。RGC は網膜上に受容野をもつ閾素子であり、これらの出力の総和(全活動度)が心理的閾値を越えると光が検出されると仮定する。したがって、TAMIT には 2 つの閾パラメータが存在する。以後、RGC の閾値を θ_n 、心理的閾値を θ_p と書くことにする。TAMIT には表 1 に示すように 8 個のパラメータが存在する。以下でこれらのパラメータについて説明する。



TAMIT

図 1. TAMIT の構成

Maximum center sensitivity	A
Maximum surround sensitivity	AQ
Radius of receptive field center	σ_e
Radius of receptive field surround	σ_i
Threshold of one neuron	θ_n
Psychological threshold	θ_p
Receptive field density	D_r
Retinal eccentricity	E

表1. モデルのパラメータ

(1) RGCのモデル

各RGCの受容野はDOG (Difference of Gaussian)関数で記述することにし、それぞれのガウス関数の標準偏差を σ_e 、 σ_i 、ガウス関数のピーク値をA、AQとする。したがって、RGCの受容野の感度分布 $G(r)$ は

$$G(r) = A (\exp(-r^2 / \sigma_e^2) - Q \cdot \exp(-r^2 / \sigma_i^2)) \quad (1)$$

である。ここに r は受容野の中心からの距離である。

既に述べたように受容野中心部の直径は網膜偏心率Eと直線関係にある⁽¹⁾、⁽²⁾。ここでは被験者 H.M. に関するデータを採用することにする。被験者 H.M. の場合は、

$$2\sigma_e = 0.55E + 4.15 \quad (2)$$

ここで Eの単位は度であり、 σ_e の単位は分である。また、受容野周辺部の直径は、網膜偏心率によらず受容野中心部の直径のほぼ3倍なので⁽²⁾、

$$\sigma_i = 3\sigma_e \quad (3)$$

とする。また、Aの値は Linsenmeier ら⁽⁴⁾の生理学的データから

$$A = 2.13 / \sigma_e^{1.23} \quad (4)$$

とした。

(2) RGCの受容野密度

中心窩付近では神経節細胞の細胞体はその受光部からずれているので、神経節細胞の細胞体密度 (個/deg²) は、錐体の密度のように網膜部位の情報処理能力と必

ずしも対応しない。Drasdo⁽⁵⁾ は、細胞体のデータから、網膜情報処理能力を示す神経節細胞の受容野密度 D_r (1平方度deg² あたりの受容野の数) を推測する式を計算した。それによると $\sqrt{D_r}$ と偏心率 E (deg) の関係は

$$\frac{1}{\sqrt{D_r}} = 0.0055(1+0.59E) \quad (5)$$

で近似できる。(5)式を積分すると 1.5×10^6 となりほぼ全神経節細胞数と一致する。この受容野密度には、X型、Y型、W型のすべての細胞が含まれている。

(3) Y型RGCの受容野密度

我々がパルス刺激を用いて測定した受容野(2)式は、いくつかの理由でY型RGCの受容野であると考えられる。そこでTAMITではRGC全体の受容野密度 D_r ではなく、Y細胞の受容野密度 D_y を考慮せねばならない。この点に関しPerryら⁽⁶⁾ は、HRP注入法によって、サルY型RGCの細胞密度をしらべた結果、9.4°以上の網膜偏心率においては偏心率によらずRGC全体の約10%であることを見い出している。網膜周辺部では細胞密度は受容野密度に一致している。それゆえ周辺部のY型RGCの受容野密度 D_y は上記のDrasdoの式から求めた D_r の10%であるとする。

6.4 並列処理計算機NCUBEにおけるシミュレーション

計算機シミュレーションでは、各細胞の受容野が同心円上に分布しているものと仮定して行った。同心円の間隔 Δr は各位置における受容野密度の逆数に等しいとした。すなわち、

$$\Delta r(E) = 1/\sqrt{D_r(E)} \quad (6)$$

ここに、 E は網膜偏心率である。また、各受容野はほぼ Δr で同心円上に等間隔に位置すると仮定した。座標 (x_0, y_0) にある細胞の出力 $O(x_0, y_0)$ は、

$$O(x_0, y_0) = \varphi \left\{ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} S(x, y) \cdot G(x-x_0, y-y_0) dx dy - \theta_n \right\} \quad (7)$$

で表わすことができる。ここに $S(x, y)$ は刺激パターンの輝度、 $G(x, y)$ は式(1)で表わされる受容野の感度分布である ($r = \sqrt{x^2 + y^2}$)。

また、 φ は次式で与えられる非線形関数である。

$$\varphi(x) = \begin{cases} x & (x \geq 0) \\ 0 & (x < 0) \end{cases} \quad (8)$$

次にすべてのY型RGCの出力の総和 G_0 を計算した。すなわち、

$$G_0 = \sum_{x,y=-\infty}^{\infty} O(x,y) \quad (9)$$

ここに $O(x,y)$ は各細胞の出力値である。

最後に全活動度 G_0 を心理的閾値 θ_p と比較し、 $G_0 \geq \theta_p$ のときに光が検出されると仮定した。本シミュレーションでは、スポット光に対する閾輝度の決定を行なったので、閾輝度での $S(x,y)$ は直径 d 、高さ ΔI の円柱関数になる。したがって、シミュレーションでは

$$G_0 = \theta_p \quad (10)$$

となる ΔI を計算することになる。すなわち

$$\theta_p = \sum_{x_0, y_0=-\infty}^{\infty} \varphi \left\{ \Delta I \cdot \int \int_{(x,y) \in S} G(x-x_0, y-y_0) dx dy - \theta_n \right\} \quad (11)$$

なる ΔI を求めることが目的である。特に $\theta_n = 0$ のときは、

$$\theta_p = \sum_{x_0, y_0=-\infty}^{\infty} \Delta I \cdot \varphi \left\{ \int \int_{(x,y) \in S} G(x-x_0, y-y_0) dx dy \right\} \quad (12)$$

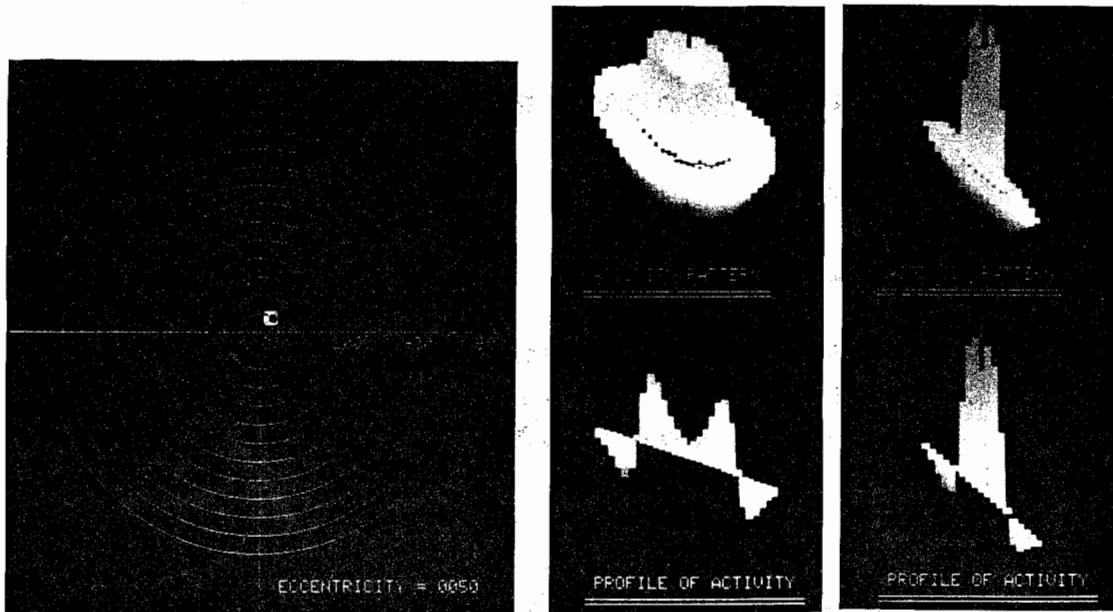
よって

$$\Delta I = \theta_p / \left[\sum_{x_0, y_0=-\infty}^{\infty} \varphi \left\{ \int \int_{(x,y) \in S} G(x-x_0, y-y_0) dx dy \right\} \right] \quad (13)$$

となる。

今回構成した網膜モデルでは、データの並列性や計算の複雑さからたった一つのスポット刺激に対するRGC群の反応を求めるのに、従来の計算機処理で数時間を費やすことになる。従って、網膜上のいろいろな位置に提示されたスポット光についてその反応特性を調べるような会話的シミュレーションが実行しにくかった。

本研究では、ハイパーキューブ型の並列処理計算機NCUBEを用い、一つのスポット光に対するRGC群の反応を256個のプロセッサに振り分けて処理した。実際には、網膜中心窩からの偏心度によって描かれる同心円上のRGC群の反応を同一のプロセッサで処理するという方法を採用した。すなわち、一つのスポット光で引き起こされるRGC群全ての反応は、偏心度をパラメータにして振り分けた複数のプロセッサで並列的に処理された。図2にシミュレーションの結果の一部を示す。図2はそれぞれ偏心度 $50'$ 、 $600'$ にスポット光が提示されたときのRGC群の反応である。



(a) 偏心度 50' (b) 偏心度 600'

上図の左の同心円は視野を示しカーソルでスポット光の提示位置を指定する（上図左には偏心度 50' を指定した時のディスプレイ画面を示す）。右上のパターンはスポット光に対するすべての Y 型 RGC の出力を、右下のパターンは断面を示す。

図 2. NCUBE によるシミュレーションの結果

6.5 モデルのパラメータの推定

表 1 に示したように TAMIT は 8 個のパラメータから成る。σ₀ と σ₁ は (2) 式と (3) 式で与えられる。A は (4) 式で与えられている。心理的閾値は同一被験者の場合、網膜偏心度 E によらず一定である。また、Q、θ_n も E によらず一定であると仮定した。D_r は既に述べたように (5) 式から計算できるが、まず E=10(deg) においては (5) 式から得られる D_r の 10% が Y であるとした。すなわち、

$$D_Y = 0.1 \cdot D_r \quad (E=10) \quad (14)$$

まず、中心窩すなわち E=0(deg) において、付近の RGC の受容野密度が一定であると仮定しその密度を変えて閾値-面積曲線の変化を調べた。結果を図 3 に示す。

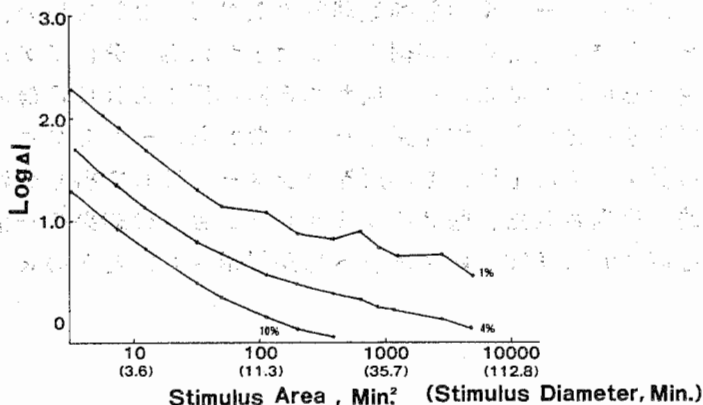


図 3. 受容野密度による閾値-面積曲線の変化

RGCの受容野密度は、閾値-面積曲線を上下に移動させる効果をもつ。次に、 $D_Y = 0.04 \cdot D_r$ として、すなわち、Y型受容野が全体の4%であると仮定して、RGCの閾値の効果調べた(図4)。

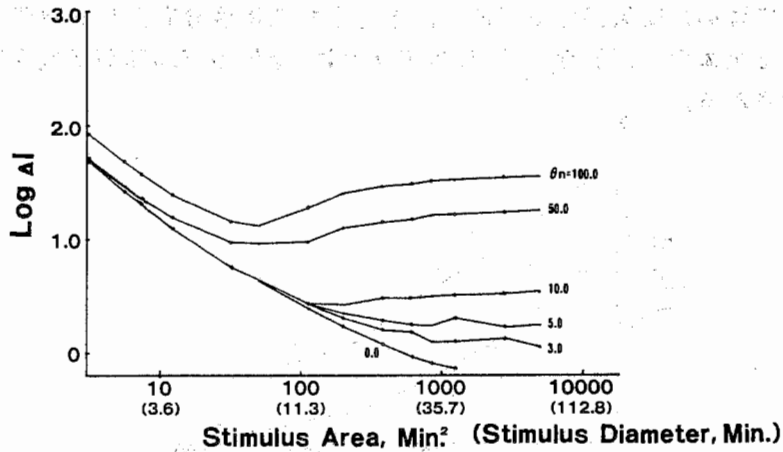


図4. 閾値-面積曲線に対するRGCの閾値の効果

θ_n が小さいときは、刺激光の直径が12'以上で閾値が上昇する。また抑制の強さ Q を大きくすると、刺激光の直径が4'以上で、閾値が上昇する(図5)。

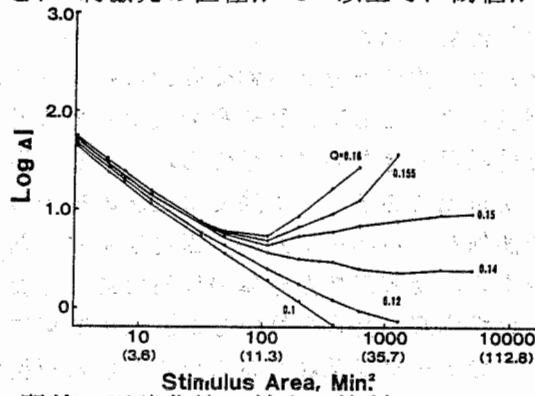


図5. 閾値-面積曲線に対する抑制パラメータの効果

被験者 H.M. の閾値-面積曲線から以下のようなパラメータの値が推定された。Y型RGC受容野に割合を表2に示す。

Eccentricity	0°	1°	2°	4°	5°	10°
Percentage of Y-cells	4	8	9	10	14	10

表2. 推定されたパラメータ値

Y型受容野の割合は、網膜偏心率が 3° 以上であれば、約10%で一定であるが、それ以下の偏心率ではその割合は減少し中心窩では4%であった。その他のパラメータ値は、 $\theta_n=2.0$, $\theta_p=480$, $Q=0.12$ であった。これらのパラメータ値によってデータにあてはめた結果が図6である。推定された密度分布から網膜偏心率が 3° 以内であれば全体の8.8%が、 1° 以内であれば全体の8.0%がY型受容野であることが予想される。

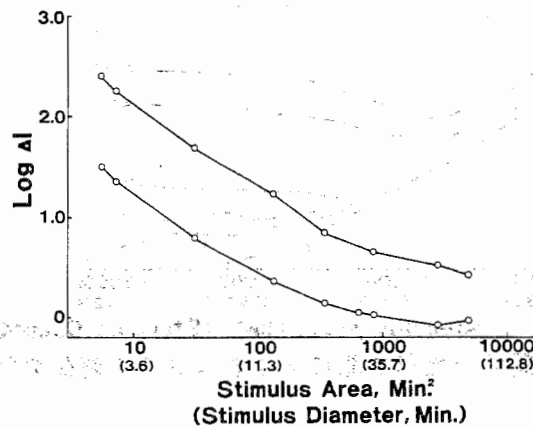


図6. TAMITによる閾値-面積曲線の予測値

6.6 X型RGCの受容野密度と視力

最小分離閾 (MAR) は、受容器である錐体間距離によっては説明できない。また順応水準や視標の輝度によってもMARや視力 (MARの逆数) が変化するが、これも錐体のレベルでは説明できないことが知られている。視力のさまざまな側面は、RGCの水準で説明可能であろう。本節では上記のモデルから推定されるX型RGCの密度と視力の関係をしらべモデルの妥当性を示す。前節では、偏心率 10° の位置でのY型チャンネルの受容野数が全体の10%であると仮定して偏心率 10° 以下でのY型チャンネルの受容野の割合を推定した。この結果に基づき、どの偏心率においてもPerryら⁽⁶⁾のデータが示唆するようにY型とW型が同数であるとして、推定されたX型RGCの受容野の密度を表3に示す。

Estimated receptive field density

Eccentricity	0°	1°	2°	3°	4°	5°	10°
Y-cell	0.37	0.29	0.17	0.12	0.08	0.08	0.019
X-cell	8.44	3.05	1.59	0.96	0.65	0.43	0.15
total	9.18	3.63	1.93	1.20	0.81	0.59	0.19

(min^{-2})

表3. 推定された受容野密度

X型RGCの受容野密度 D_x は、信号の標本化密度なのでShannonの標本化定理を用いて処理できる遮断周波数 f_c が計算できる。すなわちX型RGCが六方格子状に配列されているとすれば、RGC間の距離 $\Delta\phi$ は、

$$\Delta\phi = (2/(\sqrt{3}D_x))^{-2} \quad (15)$$

で与えられる。このとき、遮断周波数 f_c は

$$f_c = (\sqrt{3} \cdot \Delta\phi)^{-1} \quad (16)$$

である。しかし、X型RGCには、ON-中心型(正の刺激に反応)とOFF-中心型(負の刺激に反応)があるので、上記の D_x をさらに1/2倍した条件(ON-OFF Separate)でも f_c を算出した。次に遮断周波数 f_c から視力を次式によって算出し、干渉縞視力(interference fringe acuity)のデータ⁽⁷⁾と比較した(図7)。

$V.C. = 2f_c$

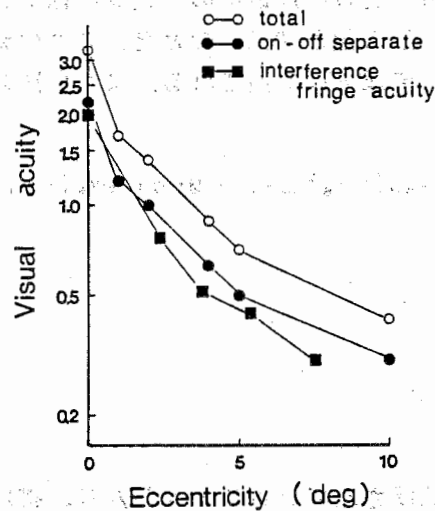


図7. 干渉縞視力とモデルの予測値

その結果、ON-OFF separate 条件で算出した視力 $V.C.$ と干渉縞視力がよく一致することが明らかにされた。ON中心型RGCとOFF中心型RGCとの対で正負の信号が検出されるのでON-OFF separate 条件が信号検出の意味で合理的である。図8は輝度を変化させて測定した視力⁽⁸⁾とモデルから予測される視力(黒三角)を示している。以上の予測値は視力の最高の値であり、輝度変化による視力低下は主として受容野の帯域通過特性が変化することによるものと考えられる。

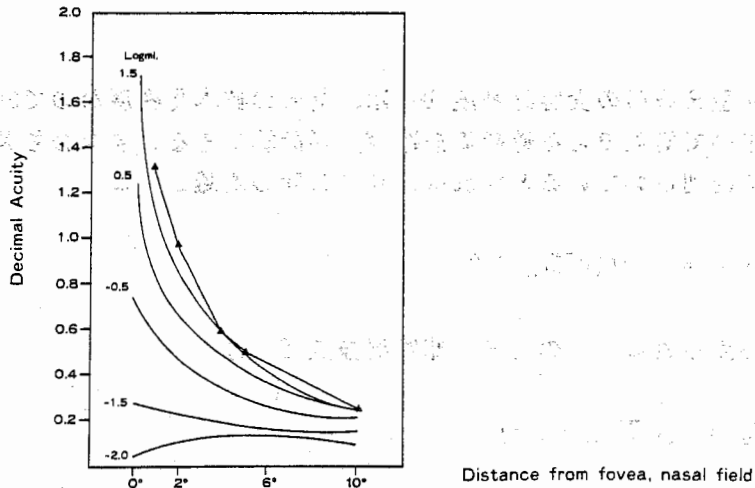


図 8. 視力の実測値 (曲線) と TAMIT と予測値 (黒三角)

6.7 考察

本研究では、網膜モデル TAMIT を通して、心理物理学的数据と単一神経細胞の出力との関係が詳細に検討された。すなわち、心理物理学に得られた閾値は、Y型 RGC の全活動度によって決定されることが示唆された。

つぎに推定された抑制の強さ Q の値は 0.12 であった。関数 $\exp(-r^2/\sigma^2)$ の体積は σ^2 に比例する。一方、受容野周辺部の半径 σ_1 は受容野中心部の半径 σ_0 の 3 倍であった。それゆえ、 Q の値が 0.11 のとき、2つのガウス関数は同一の体積となり入力刺激に対する直流成分は遮断される。また(1)式を Fourier 変換すると

$$g(\omega) = \frac{\sqrt{\pi A}}{2} (\sigma_0 \cdot \exp(-\sigma_0^2 \omega^2) - Q \cdot \sigma_1 \cdot \exp(-\sigma_1^2 \omega^2))$$

となり

$$g(0) = \frac{\sqrt{\pi A}}{2} (\sigma_0 - Q \cdot \sigma_1)$$

なのでやはり直流成分が遮断されていることがわかり、帯域通過チャンネルになっていることが理解される。網膜の情報処理能力を理解するためには、RGC 受容野特性と受容野密度の両方を明らかにする必要がある。受容野密度は信号の標本化頻度 (sampling frequency) に対応する。人間の RGC の受容野密度は、既に述べたように Drasdo⁽⁵⁾ により理論的に予測されている。しかし、X、Y、W のそれぞれについての密度分布は明かではない。最近、比較的周辺部において Y 型細胞が RGC 細胞の 10% を占めていることが明かにされた (Perry ら⁽⁶⁾; Fukuda ら、準備中) が、より中心部での密度分布はいまだ不明である。さらに、中心部では細胞体の位置と受容野の位置が異なるため、受容野密度を解剖学的に知ることは不可能である。我々は心理物理学の実験データに網膜モデル TAMIT を適用することによって、Y 型 RGC の受容野密度を推定した。その結果、網膜偏心度が 3° 以内の中心部では Y

型の割合は低下し、平均すると8.8%であった。Fukuda(準備中)は、中心部に受容野をもつRGCの軸索の太さからおよそ7%がY型であると推定している。またモデルから推定されるX型RGCの受容野密度から推定される中心外視力がこれまでに報告されているデータときわめてよく一致することが明らかにされた。これは我々が提案した網膜モデルTAMITの妥当性を実証しているものと考えられる。

6.8 文献

- (1) Inui, T., Mimura, O. & Kani, K.: Retinal sensitivity and spatial summation in the foveal and parafoveal region, *Journal of the Optical Society of America*, 71, 151-154(1981)
- (2) 乾敏郎、三村治、可児一孝: 点刺激による空間加重領域ならびに抑制領域の検討: 網膜偏心度との関係, *日本基礎心理学会誌*, 第1巻, pp. 77-84(1982)
- (3) Okamoto, Y., Mimura, O., Kani, K., Inui, T.: Characteristics of two systems of human vision using fundus perimetry. In: E. L. Gleeve(Ed.) *Documenta Ophthalmologica Proceeding Series*, Dr. W. Junk Publishers(1986)
- (4) Linsenmeier, R. A., Frishman, L. J., Jakiela, H. G., Enroth-Cugell, C.: Receptive field properties of X and Y cells in the cat retina derived from contrast sensitivity measurement, *Vision Research*, 22, pp. 1173-1183(1982)
- (5) Drasdo, N.: The neural representation of visual space. *Nature*, 266, pp. 554-556(1977)
- (6) Perry, V. H., Oehler, R., Cowey, A.: Retina ganglion cells that project to dorsal lateral geniculate nucleus in the macaque monkey, *Neuroscience*, 12, pp. 1101-1123(1984)
- (7) Campbell, E. W., Green, D. G.: Optical and retinal factors affecting visual resolution, *J. Physiol.*, 181, pp. 576-593(1965)
- (8) Slon, L. L.: The photopic acuity-luminance function with special reference to parafoveal vision, *Vision Res.*, 8, pp. 901-911(1968)

7. エネルギー最小化による画像復元とエネルギー学習

7.1 はじめに

最近、神経回路網モデルを現実的な情報処理問題に応用する試みが注目されている。このきっかけとなったのは、Hopfield & Tank⁽¹⁾の巡回セールスマン問題の研究であった。この論文で示されたように、複雑な非線形拘束条件の下である評価関数（エネルギー）を最小（大）化する最適化問題は、問題のサイズが大きくなったとき、神経回路モデルの並列構造に問題の巨大なサイズをうまくマッピングできれば、神経回路が得意とする情報処理課題の一つとなる。画像処理や視覚情報処理の分野でも、いくつかの問題が大規模な最適化問題とみなせることが指摘され、視覚情報の局所並列性を生かした、並列アルゴリズムや、神経回路モデルが提案されてきた。しかしこれらの具体的問題において、評価関数（エネルギー）やそれに対応する神経回路モデルの構造とシナプス結合荷重は、天下りの的に実験者によって与えられてきた。これは脳のモデルとしても不自然であるし、工学的な応用においても大きな障害となる。

宇野、川人、鈴木⁽²⁾は、ヒト上肢多関節運動における軌道生成の問題が、ある評価関数を最小化する問題であることを示した。さらに我々^{(3)、(4)}は、この評価関数を運動の例から学習によって獲得し、しかも回路の並列演算によってこの評価関数を最小化できる4層神経回路網モデルを提案した。

本章では、画像復元や初期視覚の問題について、軌道生成の神経回路モデルと同様に最小化すべきエネルギーを画像の例から学習によって獲得し、さらに回路の並列演算でそのエネルギーを最小化する神経回路モデルを提案する。

7.2 初期視覚・画像復元と神経回路モデル

7.2.1 初期視覚

ヒトの視覚情報処理の第一段階は初期視覚と呼ばれており、その目的は、網膜像（2次元画像）から実世界（3次元物体）の様子を推測することである。すなわち初期視覚とは、2次元の原画像の強度を決定している4つの要因、(1)可視表面の幾何学的構造、(2)可視表面の反射率、(3)光景に対する照明、(4)観察点を、2次元画像情報をもとに分離表現することである⁽⁵⁾。Poggioら⁽⁶⁾は、この過程が、3次元物体から2次元の濃淡画像への写像である光学のちょうど逆になっているので、初期視覚を逆光学と呼び、それが数学的には不良設定問題であることを示している。従って画像データ以外にも何らかの拘束条件がなければ、(i)解が存在しない、(ii)解が一意に定まらない、(iii)解がデータに対して連続とならない、のうちいずれかが生じてしまう。

Poggio & Koch⁽⁷⁾は適当な拘束条件を与えて、結果として得られる最適化問題を標準正則化理論を用いて解くとともに、並列神経回路モデルも提案した。さて標準正則化理論では評価関数は二次形式として与えられる。従って標準正則化理論によって提供される作用素は線形であり、並列神経回路モデルも線形のものに限られる。

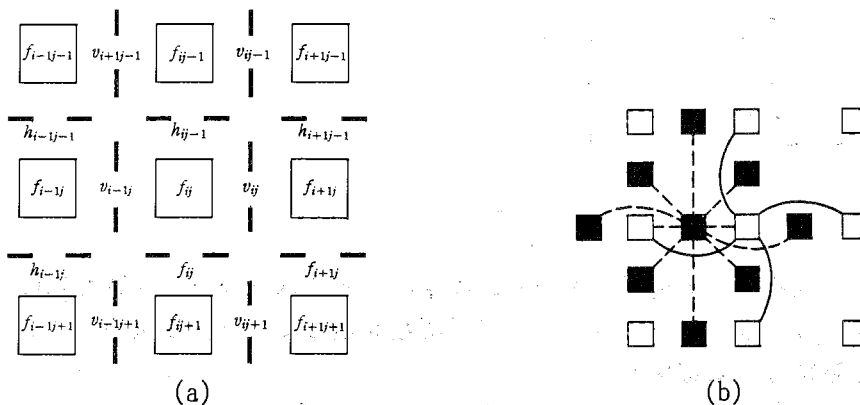
このような理論でも、心理物理学のいくつかの現象を再現したり、コンピュータビジョンのために有効なアルゴリズムを与えることはできるが、画像に不連続性を含む本質的に非線形の問題を取り扱うことはできない。

7.2.2 マルコフ確率場とライン過程

Geman & Geman⁽⁸⁾は、1984年に少なくとも4つの点で画期的な進歩をもたらした論文を発表した。彼らは、初期視覚の問題ではなく、ノイズに汚されたデータから原画像を推定し復元する画像復元の問題を取り扱ったが、その基本的なアイデアが初期視覚の問題にも適用できることは、この論文の後を追う MIT や Brown 大学グループの数々の論文^{(6), (9), (10), (11)}からも明かであろう。彼らは第1に画像のマルコフ確率場モデルが Gibbs 分布のエネルギーを定義することと等価であることを示した。第2に、ライン過程と呼ばれる画像の不連続性を表現する隠れた(仮想の)確率変数を導入した。第3に画像復元の問題を、最大事後確率を与える原画像を推定することととらえて、対応するエネルギー最小化が局所並列演算(確率緩和)によって実行可能であることを示した。第4にこの確率緩和に温度のゆっくりとした低下(シミュレートされた焼きなまし: Simulated Annealing)を導入することにより、エネルギー最小化が厳密に達成できることを示した。

7.2.3 神経回路モデルによる奥行き面の再構成

Kochら⁽⁹⁾は、Geman & Geman⁽⁸⁾のライン過程とマルコフ確率場の概念を応用し、Hopfield型の神経回路で、不連続性を本質的に含む初期視覚の問題を解いた。この神経回路は、両眼視によって雑音を伴う奥行きデータが格子状の計測点上にまばらに与えられているとき、滑らかな曲面を再構成する。物体の端に相当する部分では、奥行きデータが不連続になる。この不連続を誤って曲面化してしまわないように、ライン過程と呼ばれる隠れた確率変数を格子点の間に配置する。図1(a)の正方形が各画素でのデータ、その間にある線がライン過程を示す。



- (a) 格子状に配列した奥行きデータと不連続を表すためのライン過程
- (b) 奥行き(白四角)とライン過程(黒四角)を表すニューロン間の相互作用

図1. 初期視覚の問題を解決する神経回路モデルの構造

Geman & Geman は「焼きなまし(Simulated Annealing)」による確率緩和でエネルギー最小化を達成したが、Koch らは神経回路の並列演算でエネルギーの最小化を達成した。これは計算時間の点で確率緩和よりはるかに優れている。

図 1 に示す神経回路モデルを簡単に説明しよう。格子点上での奥行きデータの推定値を f 、観測により得られた奥行きデータを d 、格子点の間に配置するライン過程で水平に並んだ奥行きデータの不連続を示すものを v 、垂直に並んだデータの不連続を示すものを h とする。 f は実数、 h, v は、0 と 1 の間の数とし、1 のとき、完全な不連続を、0 のときには完全な連続を表している。変数の定義:(1)式、エネルギーの各項の定義:(2)-(5)式、及び変数の時間変化則:(6)式を以下に示す。

$$h_{ij} = g(m_{ij}) = \frac{1}{1 + \exp(-2\lambda m_{ij})} \quad (1)$$

$$E_1 + E_D = \sum_{ij} (f_{i,j+1} - f_{ij})^2 (1 - h_{ij}) + C_D \sum_{ij} (f_{ij} - d_{ij})^2 \quad (2)$$

$$E_L = C_U \sum_{ij} h_{ij} (1 - h_{ij}) + C_P \sum_{ij} h_{ij} h_{i,j+1} + C_C \sum_{ij} h_{ij} + C_L \sum_{ij} h_{ij} [(1 - h_{i+1,j} - v_{ij} - v_{i,j+1})^2 + (1 - h_{i-1,j} - v_{i-1,j} - v_{i-1,j+1})^2] \quad (3)$$

$$E_G = C_G \sum_{ij} \int_0^{h_{ij}} g_{ij-1}(h_{ij}) dh_{ij}, \quad g(U_i) = \frac{1}{1 + \exp(-2\lambda U_i)} \quad (4)$$

$$E = E_1 + E_D + E_L + E_G \quad (5)$$

$$\left. \begin{aligned} \frac{df_{ij}}{dt} &= - \frac{\partial E}{\partial f_{ij}} \\ \frac{dm_{ij}}{dt} &= - \frac{\partial E}{\partial h_{ij}} \end{aligned} \right\} \quad (6)$$

ここで、 i, j は格子点の位置を表わす。簡単のために水平なライン過程に関する式のみを示す。 m_{ij} は、 g というシグモイド関数を通して、 h_{ij} を決める。これはニューロンの膜電位に対応する。

エネルギーの各項の意味は次の通りである。 E_1 は奥行きデータの連続性、 E_D は計測されたデータの信頼性、 E_L は不連続(ライン過程)の生じ方、 E_G は m が発散しないことをそれぞれ要求する。これらのエネルギー関数の各項の係数 $C_D, C_U, C_P, C_C, C_L, C_G$ は、それぞれの項の重みを決定する係数である。次節で説明する NC

UBE上の実験では、Kochらの用いたパラメータとほぼ同じ値を用いた。これらの係数を決定するのはなかなか困難である。

この神経回路モデルのシミュレーションには、 N 個の格子点に対して、 $3N$ 個のニューロンの状態変化を表す複雑な非線形常微分方程式を解く必要がある。図1(b)に奥行きを表わす格子点(白ヌキ四角)と、ライン過程を表わす格子点(黒四角)がどのように相互作用するかを示してある。これからわかるように1つのライン過程に対応する1個のニューロンは、周囲の10個のニューロンと相互作用する必要があり、格子状に配列された画像処理専用の並列計算機では結合が足りなくなることがわかる。

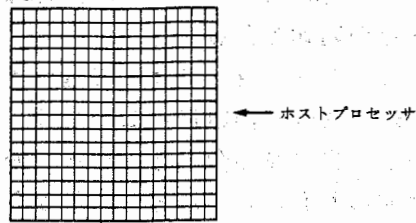
7.3 NCUBE上での神経回路モデルによる画像復元

本節では256個のノードを持つハイパーキューブ構造の強並列計算機NCUBE/10⁽¹²⁾,⁽¹³⁾に前節の神経回路モデルをインプリメントし、画像復元の問題を解く。NCUBE上のプログラミングの詳細は池田ら⁽¹⁴⁾を参照されたい。

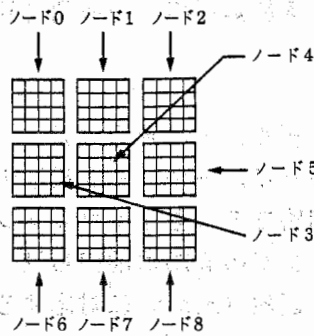
まず、2次元に配列している1画素(m, h, v に対応する3個のニューロン)の処理を、おのおの1ノードに処理させるべくプログラムを行なったが(図2b)、この方法では、NCUBEの能力がうまく引き出せなかった(表1参照)。原因は、二次元格子状に配列している画素の処理を、神経回路モデルのニューロンの昇順そのままにノードに割り付けたことにある。それぞれのノードは、受け持ちの画素と隣接する8つの画素を受け持つ8つのノードと密に情報交換する必要があるが、この通信が、ハイパーキューブ上で直接隣接しない2ノード間で頻繁に行なわれるため、ノードは、処理の大半の時間を他のノード間の通信の中継処理や、通信に必要なルートの空き時間待ちに費やしてしまい、計算速度が上がらないことがわかった。

分割方法	データ	実行時間	試行回数	相対速度
分割なし	32 * 32	12 秒	100	1.0
2次元	32 * 32	19 秒	100	0.63
1次元	256 * 256	91 秒	600	506

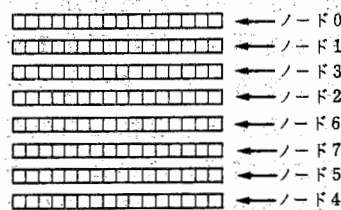
表1. NCUBE上の神経回路モデルによる画像の復元実験のパフォーマンス比較



(a) 分割なし



(b) 2次元分割



(c) 1次元分割

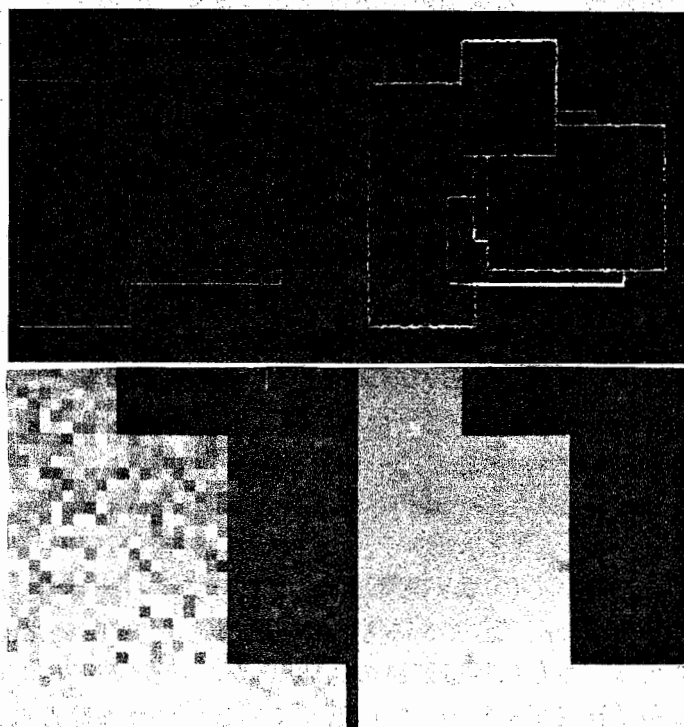
- (a) すべての画素に対応するニューロンをホストで処理
 - (b) 近傍の画素に対応するニューロンをまとめて1つのノードで処理 (2次元分割)
 - (c) 横1列の画素に対応するニューロンをまとめて1つのノードで処理 (1次元分割)
- 図2. 神経回路モデルのニューロンのNCUBEのノードへの割り付けの様子

そこで、ハイパーキューブのリンクにうまく適合するようにモデルを移植するために問題の並列化から再度検討し直し、2次元に配列した画素の横方向の列をまとめてノードに割り当てた(図2(c))。このように分割すれば、隣接している画素の左右2つの画素は、同一のノードに含まれ、左上、真上、右上の画素は、上隣のノードに、左下、真下、右下の画素は下隣のノードに含まれる。そのため、密に通信するのは、上下のノードのみであることがわかる。つまり、ノードを1次元にリンクすれば、特定のノード間で行なわれる通信が他の通信の終了を待つことはない。このリンクは言わば、ハイパーキューブに沿った一筆書きであり、それはまた双方向にデータが流れるパイプラインを形成している。このリンクの構造は幾何学的で、ハイパーキューブの次元が大きくなると複雑化する。今回は、計算機プログラムでリンクを決定しそれを利用した。ノード割り付けの様子を図2下に示す。

今回採用した1次元鎖の拡張として、その両端が互いに接続している1次環、1

次鎖を平行に配置して鎖のなかの相対するノードをさらに鎖状に接続してできるメッシュ状のもの（2次鎖）、1次の場合と同様にして得られる2次環、ひいては、これらをn次にまで拡張した形態も、ハイパーキューブ構造の内部にノードを無駄にすることなくリンクできる。特に2次鎖構造は現在研究しているモデルのシナプス結合荷重の学習や、ライン過程を介して異種視覚情報を統合する場合に適する。

図3に示すように6つの濃淡値を取る幾何学的な人工画像(256*256)を考える。これに標準偏差約20のホワイトな正規乱数を加える。画像の最大濃淡差は150であった。わずか50ステップ程度(8秒の実行時間)で輪郭線を含めて原画像を再現できた(図3上、下)。



上段：ライン過程の状態変化。左上：(6)式に従う1回の更新の後の状態
 右上：200回の更新の後の状態
 下段：濃淡値の状態変化。左下：ノイズに汚されたデータ
 右下：(6)式に従う200回の更新の後の状態
 ただし全画像の中心部分のみ示した。

図3. 雑音をともなう画像データからの不連続を考慮にいれた復元実験

計算時間については表1に示すようにホスト単独の約500倍の計算能力が発揮された。ホストの能力が0.5MIPSと低めに見積ってもこれは250MIPSに相当し、ほぼNCUBEの最大能力が引き出せるように、モデルをハイパーキューブ上で処理することができた。神経回路モデルと、並列計算機でのシミュレーションとの両面から満足できる結果がえられたといえよう。

7.4 エネルギー最小化とエネルギー学習を同時に行なう神経回路モデル

前節の実験で、原画像に加える正規乱数の分散を増やしていくと、本来画像の輪郭のない場所にライン過程が立つ ($h_{ij}, v_{ij} = 1$) ようになる (図3上のノイズのように見える過剰なライン過程に注意)。計測データの信頼性が下がれば当然(2)式の C_0 を減らさねばならないし、不連続の少ない画像データに対しては(3)式の E_L を増やさねばならない。しかし これを実験者がいちいち決めていたのでは何のための画像復元かということになってしまう。(2)-(4)式に現われるパラメータあるいはエネルギーの形は、画像のモデルあるいは物理世界の拘束条件として導入されたものなのだから、ある決められた画像のセットに対して一意に決まるはずである。エネルギーを学習によって神経回路モデル内に獲得できないだろうか。

f がニューロンの状態、 d が画像復元の計測データを、 C がエネルギーを特徴づけるパラメータを表わすベクトルとする。また f^* は画像の真の状態を示すとする。(1)-(6)式で与えられる神経回路モデルは、エネルギーを $E(f, d, C)$ と表わせば、ほぼ E の最急降下方法に状態変化を起こし、画像の推定値 f^* として、

$$\frac{\partial E(f^*, d, C)}{\partial f} = 0 \quad (7)$$

を満たすものを与える。今画像の推定値と真の値の二乗誤差 $(f^* - f^{\wedge})^2$ を最小にするように、エネルギーを定義するパラメータ C を変化させることを考える。

$$\frac{\partial C}{\partial t} = - \frac{\partial (f^* - f^{\wedge})^2}{\partial C} \quad (8)$$

まず陰関数の定理を用いて(7)式から $\partial f^* / \partial C$ を計算する。 f^* で E が極小であることと、 E の主要項が E_D, E_G であるという仮定から、 $(\partial^2 E(f^*, d, C) / \partial f^2)^{-1}$ を正定数で置き換える。さらに神経回路の状態 f が、パラメータ C の学習時間のスケールではほとんど安定平衡状態 f^* の近くにいるという準静近似の仮定をすれば、結局パラメータ C の変化則として次式が得られる。

$$\frac{\partial C}{\partial t} = \varepsilon \frac{\partial^2 E}{\partial f \partial C} (f - f^{\wedge}) \quad (9)$$

各ニューロンの状態遷移が次の方程式に従うことに留意すれば、

$$\frac{df}{dt} = -a(f) \frac{\partial E}{\partial f}, \quad 0 < a(f) \ll \infty \quad (10)$$

結局、エネルギー関数 C を用いて、ニューロンの状態 f の時間変化は、

$$\frac{\partial C}{\partial t} = -\varepsilon \frac{\partial}{\partial C} \left(\frac{df}{dt} \right) (f - f^*) \quad (11)$$

を得る。さらに(6)式を計算すればわかるが一般に(11)式の df/dt は次のようにパラメータ C をシナプス荷重としたシナプス後電位の線形和となることがわかる。

$$\frac{df_{ij}}{dt} = \sum_k C_{ij}^k g_k(f_{1m}, h_{1m}, d_{ij}) \quad (12)$$

$$\frac{dm_{ij}}{dt} = \sum_k C'_{ij}{}^k g_k(f_{1m}, h_{1m}, d_{ij}) \quad (13)$$

g_k, r_k は非線形関数、添字 $1m$ はエネルギー(あるいは神経回路の相互作用)が局所的であることから格子点 ij の近傍の値のみを取る。従って最終的に、エネルギーを定めるパラメータつまりシナプス結合荷重の学習則は、次式で与えられる。

$$\frac{dC_{ij}^k}{dt} = \varepsilon g_k(f_{1m}, h_{1m}, d_{ij}) (f^*_{ij} - f_{ij}) \quad (14)$$

$$\frac{dC'_{ij}{}^k}{dt} = \varepsilon r_k(f_{1m}, h_{1m}, m_{ij}) (h^*_{ij} - h_{ij}) \quad (15)$$

(12), (13)式のニューロンの状態の時間変化に比べて、(14), (15)式で与えられるシナプス結合荷重の変化が十分遅くなければいけないことをもう一度注意しておく。

さて(12)-(15)式をもう一度検討すると、これはよく知られた異シナプス可塑性学習則となっている。つまり図4に示すように、1個のニューロンは、通常のシナプス入力 g_1, \dots, g_N をシナプス荷重 C_1, \dots, C_N で重み付けて、その状態を(12), (13)式に従って変化させている。

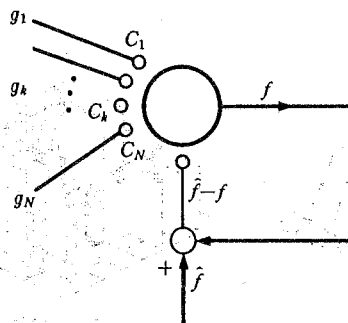
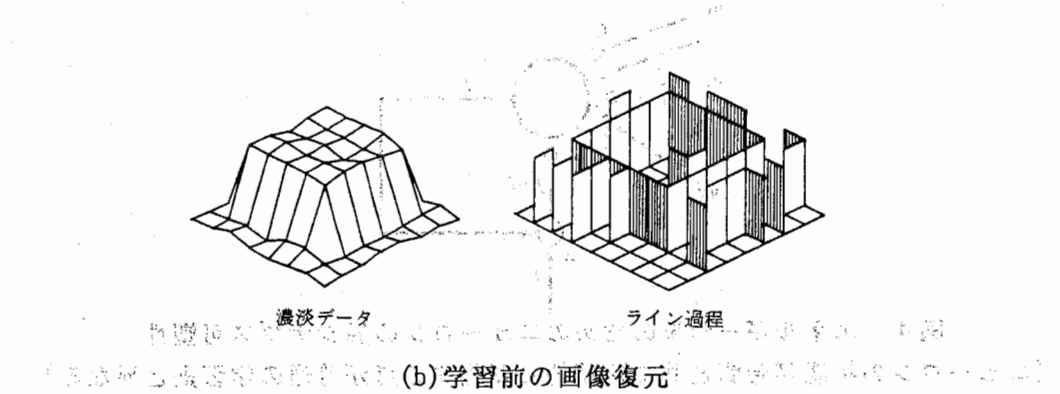
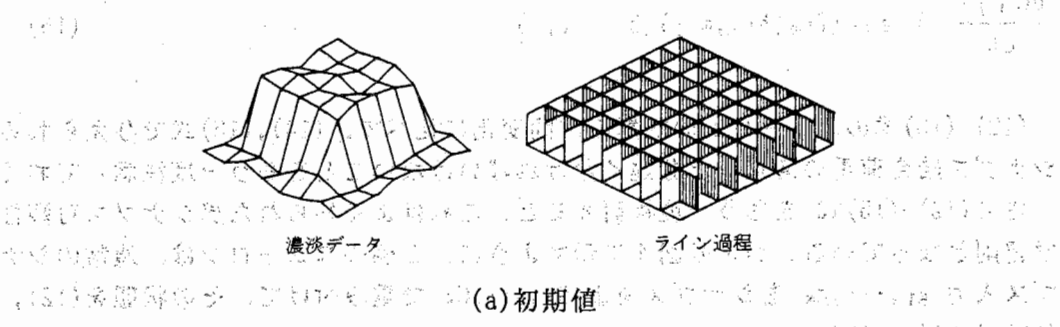
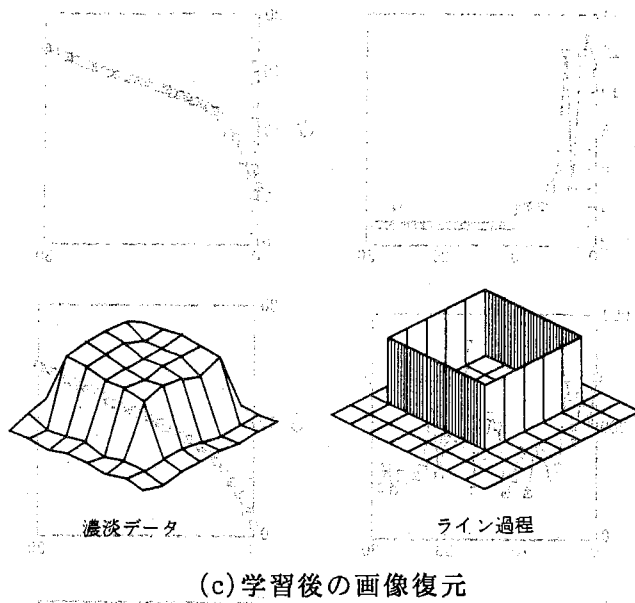


図4. エネルギー学習のためのニューロンの異シナプス可塑性
(ニューロンの状態が時間と共に連続的に変化する点が普通の学習則と異なる)

このニューロンは誤差信号 $f^* - f$ を第2の種類シナプス入力として受け取っており、2種類のシナプス入力と同時に活動したときのみ、第1種類のシナプス荷重だけが変化する。(14),(15)式では画像の各点で、エネルギーを定義するパラメータ値が異なっても良いようになっている。画像が等方的であれば、(14),(15)式を ij について平均すればよい。結局、エネルギーの最小化を行いながら(12),(13)、そのエネルギーをより画像のモデルとして適当なように学習によって変更する((14),(15))ことが一つの神経回路で達成できるのである。

我々は、この学習則を、 8×8 の格子点上に与えられた簡単な幾何学図形(5×5 の正方形:図5(a))である2次元画像(最大濃淡差 200)のエネルギーを特徴づけるパラメータの学習に応用した。まず標準偏差 12 のホワイトな正規乱数を各格子点に加える(図5(a)左)。神経回路の濃淡データ f の初期値としては、ノイズに汚されたデータをそのまま用いる。一方ライン過程の初期値はすべて 0.5 とする(図5(a)右)。エネルギーの形としては、式(2)から(5)のように選び、エネルギーを特徴付けるパラメータは、NCUBE上のシュミレーションと同じとした。図5(b)に示すように、エネルギーの学習前には原画像で不連続のない場所にまでライン過程が立ってしまい、エネルギーのパラメータの値が不適当であることが分かる。これに対して約80個の同じ統計的性質をもつ画像のパターンを用いて学習を行った後では、図5(c)に示すように、本来不連続のある場所のみライン過程が立ち濃淡データも滑らかに再構成されていることが分かる。





- (a)濃淡データの初期値（左、生データ）とライン過程の初期値（右）
 (b)学習前の神経回路モデルによる画像復元の結果
 (c)エネルギーのパラメータを学習した後の神経回路モデルによる画像復元の結果

図5. 簡単な2次元画像に対するエネルギー学習の実験

図6(a)に、標準偏差10のノイズを加えたときの30個のパターンを用いた学習中のライン過程の誤差、濃淡値の誤差、エネルギーを特徴づける6個のパラメータの変化を示す。図6(b)は、ノイズの標準偏差が小さいとき($\sigma=8$)と大きいとき($\sigma=10$)とを比較すると、真の画像に加えるノイズの標準偏差が大きいときには、 C_D は小さな値に収束することを示していて、エネルギー学習がうまく進んでいることが分かる。

ここで提案した学習則はエネルギーを特徴づけるパラメータの変化則として導出したが、図4を見れば、これはエネルギーの形そのものの学習に拡張できることがわかる。つまり、図7に示すように、入力層と出力層の二層からなる神経回路モデルがエネルギーの形そのものを学習し、さらに回路のダイナミクスにしたがって、獲得したエネルギーを最小化するモデルになっている。

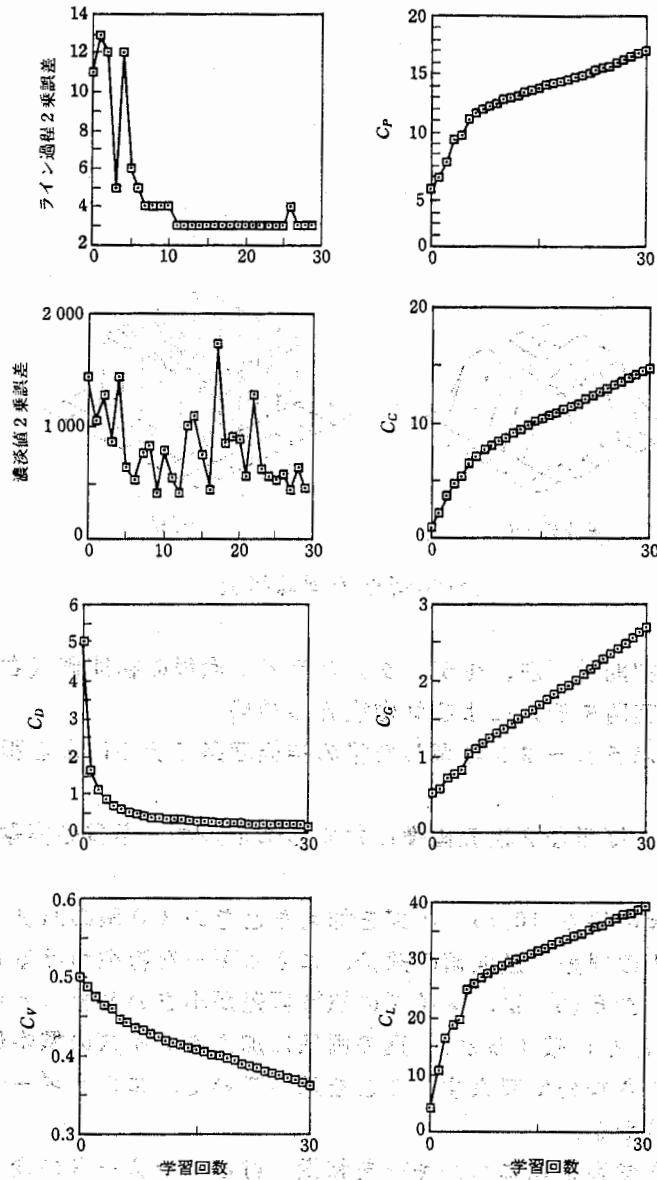


図6(a) エネルギーの学習中における、神経回路モデルが推定したライン過程自乗誤差、濃淡値の自乗誤差、エネルギーを特徴付ける6つのパラメータの学習回数に伴う時間変化

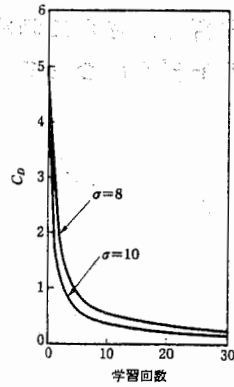
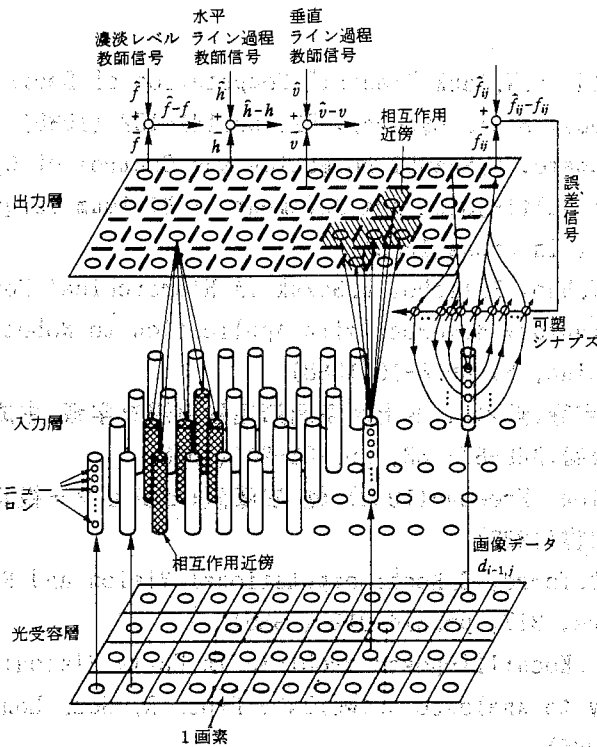


図6(b) 画像に加えるノイズの標準偏差によるパラメータ C_D の収束先の違い



2層とも、2次元画像の格子点に対応して、ニューロンがどの画素に対応するかが決まっている。入力層から出力層への結合も、出力層から入力層への結合も、画像モデル（エネルギー）の局所性から、その範囲は相互作用近傍に限られている。この神経回路モデルは、その状態をフィードバック結合によって連続的に変化させながら、シナプスの結合荷重を学習によって変えていく。

図7. エネルギーの形そのものを学習する2層神経回路モデル

7.5 おわりに

本研究では、初期視覚・画像復元問題で近来大きな進歩の見られた局所並列アルゴリズム、ライン過程を含むマルコフ確率場、神経回路モデルによるエネルギー最小化について紹介した。神経回路モデルを強並列計算機NCUBE上にインプリメントし、画像の不連続性を含めた画像復元に成功した。またエネルギーの学習と最小化を同時に行える神経回路モデルを提案し、簡単な例についてパラメータ学習がうまく行えることを確認した。

初期視覚における線形アルゴリズム（フィルター）つまり二次形式のエネルギーを表現する行列を画像データから推定する試みがMIT(10)とブラウン大学(11)で行なわれている。本論文で提案した神経回路は非線形アルゴリズム、一般的なエネルギーを学習できるという点でこれらの研究より進んでいることを最後に強調しておく。

7.6 文献

- (1) J.J.Hopfield, D.W.Tank:"Neural" Computation of Decisions in Optimization Problems, Biol. Cybern., 52, pp.141-152 (1985)
- (2) Y.Uno, M.Kawato, R.Suzuki:Formation and Control of Optimal Trajectory in Human Multi-joint Arm Movement -- Minimum Torque-Change Model, Biol. Cybern., in press(1989)
- (3) M.Kawato, Y.Uno, M.Isobe, R.Suzuki:A Hierarchical Neural Network Model for Voluntary Movement with Application to Robotics,IEEE Control Systems Magazine, 8, pp.8-16(1988)
- (4) 前田,川人,宇野,鈴木:ヒト上肢多関節運動軌道を学習,生成する多層神経回路モデル,信学技報,MBE-87-133, pp.233-240(1988)
- (5) D.Marr:Vision, Freeman(1982) 乾,安藤訳:ビジョンー視覚の計算理論と脳内表現ー,産業図書(1987)
- (6) T.Poggio, V.Torre, C.Koch:Computational Vision and Regularization Theory, Nature, 317, pp.314-319 (1985)
- (7) T.Poggio, C.Koch:"Ill-posed Problems in Early Vision: from Computational Theory to Analogue Networks", Proc. R. Soc. London, Ser.B 226, pp.303-323(1985)
- (8) S.Geman and D.Geman:Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images, IEEE Trans. Pattern Anal. Machine Intell., PAMI-6, pp.721-741(1984)
- (9) C.Koch, J.Marroquin, A.Yuille:Analog "Neural" Networks in Early Vision, Proc. Natl. Acad. Sci. USA, 83, pp.4263-4267(1986)
- (10) A.Hurlbert, T.Poggio:Synthesizing a Color Algorithm from Examples, Science, 239, pp.482-485 (1988)
- (11) D.Kersten, A.O'Toole, E.Sereno, D.Knil, J.Anderson:Associative Learning of Scene Parameters from Images, Applied Optics, 26, pp.4999-5006(1987)
- (12) 三宅:ニューラルコンピューテーション,信学会誌,70,pp.1255-1262(1987)
- (13) NCUBE:NCUBE HANDBOOK v1.1 (1986)
- (14) 池田,三宅,川人,乾,淀川,鈴木:初期視覚の計算理論と神経回路による解法,第2回生体・生理工学シンポジウム論文集,pp.137-140(1987)
- (15) 川人,池田,三宅:神経回路の学習と視覚情報処理,テレビジョン学会誌,Vol. 42, No. 9, pp.918-924(1988)

8. ネオコグニトロンによる手書き数字認識(1), (2), (3)*

8.1 はじめに

最近になって並列処理計算機という新しいタイプのマシンが計算機工学の分野で注目を浴びつつある。並列処理計算機は情報が並列的に処理されるような問題については非常に効果的に働く。ニューラルネットはそのような性質を持つ問題の一つと考えられよう。本章においては、並列処理計算機NCUBEに最も複雑なニューラルネットと言われるネオコグニトロン(4)をインプリメントした結果について述べる。NCUBEはハイパーキューブ状に結合された512個のノードプロセッサとフロントエンドのホストプロセッサ1個から成る。ノードプロセッサは2MIPS, 0.5MFLOPSの計算能力を有する。NCUBEは、コネクションマシン(5)のような「高粒度マシン」に対し、「低粒度マシン」の範中に位置づけられる。

8.2 ネオコグニトロンの性質

8.2.1 並列処理構造

図1にネオコグニトロンの構造を示す。ネオコグニトロンは入力層 U_0 と4段の処理ステージから成る。各段は2つの層(S細胞層とC細胞層)が縦続接続されて構成される。図1において、これら各層は細い線で描かれた長方形で示されている。各層は細胞面と呼ばれるいくつかのサブグループに分割されている。これら細胞面は太線で描かれた長方形によって図示されている。

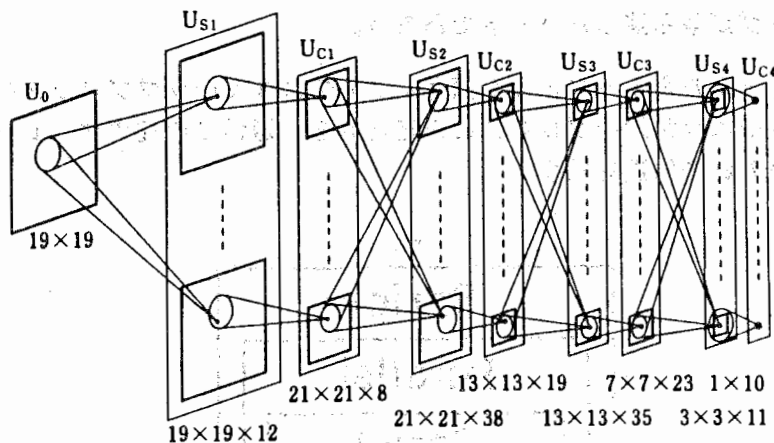


図1. ネオコグニトロンの構造

* NHKとの共同研究であり本章記述主要部分はNHK伊藤崇之氏の助力を得た。

S細胞の細胞面を「S面」、C細胞の細胞面を「C面」と呼ぶ。一つのS面は前段の多数のC面からの入力を受ける。一方、一つのC面は同一ステージ（段）の一つのS面からのみ入力を受ける。このS面とC面の対を「SCモジュール」と呼ぶ。福島らはすでにネオコグニトロン⁽⁶⁾による手書き数字認識システムを構成した。各層における総細胞数を図1の各層の下に示した。

さて並列処理計算機のプログラムにあたっては、解くべき課題をどのようにして複数のサブ問題に分割できるかを知ることが重要である。問題の分割法の一つはデータを多くのサブデータに切り分けるやり方である。この方法は各サブデータが独立に同じオペレーションで処理されるような場合にのみ有効である。もう一つの方法は処理そのものを独立に実行できる多くのサブプロセスに分割する手法である。ニューラルネットにおいては、処理を分割するのにいろいろなやり方が考えられる。以下に述べる方法はこのような分割法のいくつかの候補である。どの方法が最良であるかは採用されるハードウェアの性質に依存する。今回のインプリメントにおいて、われわれは細胞面による分割を選択した。すなわち、一つのSCモジュールがNCUBEの一つのノード（SCノード）に割り当てられる。

(1) 細胞面による分割

巨視的観点から言えば、各細胞面は前層からのデータを受け取り、後続の層にデータを送る。この過程は細胞面毎に並列に実行される。

(2) 細胞による分割

一つの細胞面において、各細胞は独立に情報を処理する。このような分割は数万個のプロセッサを持つような「高粒度マシン」に適していると言えよう。

(3) シナプスによる分割

微視的観点からは一つの細胞の各シナプスが他の細胞からの一つの入力を受け取っておりそれぞれ独立に処理されていることがわかる。このような分割にはニューロチップが最も適したハードであろう。

つまりニューラルネットのような問題については多くのサブプロセスのデータが組み合わされそれが後続のプロセスに伝達される。通常この種のオペレーションは図2のような入れ子構造を成す。

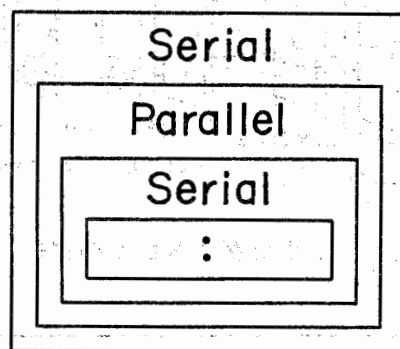


図2. 処理の入れ子構造

8.2.2 ネオコグニトロンのその他の性質

ここで並列処理に関わるネオコグニトロンのその他の性質について議論する。

- (1) 全てのS面は同一処理を実行する。S面の全細胞は同一の機能を有し、異なる点はその入力結合だけである。同じことがC面についても成り立つ。
- (2) 第1ステージの細胞面の多くは各入力パターンについてゼロでない出力を出す。一方、第2ステージ以降の奥のステージではゼロでない出力を出す細胞面がわずかになってしまう。この違いは、各ステージにおいて特徴を検出するための受容野のサイズの違いに起因する。すなわち、初段のS面は特定方位の線分を検出するが、そのような線分特徴はほとんど全ての入力パターンが共通して持っている特徴である。一方、もっと奥のステージで観測される大局的特徴は各カテゴリーについて特異的なものであり、反応する細胞の数は少なくなる。
- (3) 第1と第2のステージ間の結合は all-to-all の形態で形成される。すなわち、第2ステージのS面は第1ステージのほとんど全てのC面からデータを受け取っている。このような結合形態については、奥の層ではそれほど密集した形態にならない。つまり、一つのS面は前段の少数のC面からだけデータを受け取る。

8.3 インプリメント

8.3.1 ノードの機能

前述したように一つのSCモジュールは一つのノード(SCノード)に割り当てられる。この方法は以下のような利点を有する。

- (1) データ転送の単純化：SCモジュールの処理は転送すべきデータ量は大きい
が、ノード間のデータ交換は単純である。
- (2) 拡張性：SCノードの数は学習パターンや認識すべきパターンの変更に
応じて増やすことができる。NCUBEに余ったノードがある限り新しいSCノード
を容易に追加できる。

なお、ネオコグニトロンはV面と呼ばれる別のタイプの細胞面を持つ。V面は各ステージの一つずつ存在する。V面はもう一つ別のノードに割り当てられている。

8.3.2 データ転送

一つのSCノードは限られた数のリンクを通じて他の多くのノードからデータを受け取る。これはデータ転送における衝突を引き起こす。これを避けるために、われわれはブロードキャストイング⁽⁷⁾とギャザリングという手法を用いた。

図3に3次元ハイパーキューブにおけるブロードキャストイングの手続きを示した。nサイクルのデータ転送で、ルートノードのデータは、n次元ハイパーキューブ内の他の全てのノードに伝えられる。データを集めるギャザリングは、図3のデータの流れを逆転することによって実現する。中間のノードは受け取ったデータを

自分自身のデータと組み合わせ、それを次のノードへ転送する。このようなアルゴリズムは、全てのノードが同期して働くという拘束条件があるにもかかわらず、データの衝突を防ぐ効果がある。転送時間が大きくなるために、もしデータがゼロの要素のみで構成されている場合にはSCノードからのデータ転送は省かれる。

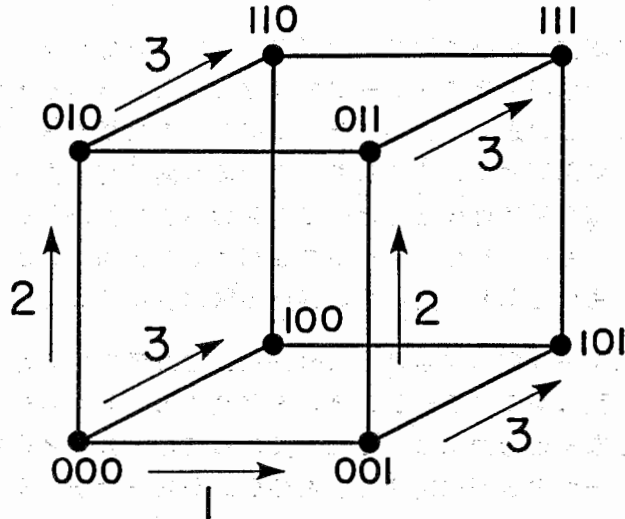
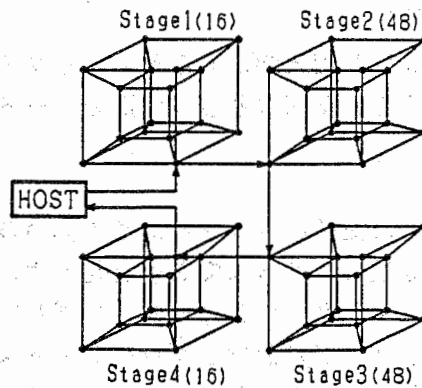


図3. 3次元ハイパーキューブにおけるブロードキャスト

8.3.3 結果

図4にノードの割り当ての概略を示した。各ステージの機能はSCモジュールの数を満足するだけのノード数を持つサブキューブによって実現される。Vノードはサブキューブのルートノードに割り当てられる。ルートノードはホストや他のVノードとのデータの送受信を行なう。図1のシステムを実現するのに128個のノードを要した。したがって、512ノードのNCUBEに4個のネオコグニトロンが同時に実行できる形でインプリメントした。



()内の数字は各サブキューブにおけるノードの数を示す

図4. NCUBEにおけるネオコグニトロン構成

表1はパターン“8”を認識するに要する計算時間の一例である。通常“8”は最も長い計算時間を必要とする。同一パターンについて Micro VAX IIでの計算時間との速度比較を表1に示した。図5に認識結果のCRT表示例を示す。

	NCUBE	microVAX II
time(sec)	0.95/4chars.	17.3
speed ratio	72.8	1.0

表1. ネオコグニトロンの実行時間

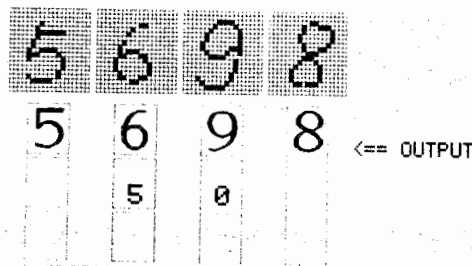


図5. 認識結果のCRT表示の一例

8.4 考察

8.4.1 データ転送

ブロードキャストイングやギャザリングは多くの結合を持つネットワークのシミュレーションに有効であるが、各ステージで全データがVノードに集中することが一つのボトルネックになっている。しかしながら、以下の評価によってこのことが無視し得ることがわかる。隣接するノード間のデータ転送時間を測定し、線形減衰評価値を算出した。すなわち、

$$(\text{転送時間}) = 1.1 \times (\text{データカウント}) + 298.1 (\mu \text{sec})$$

第1ステージからの出力データが最も大きいので第1ステージから第2ステージへのデータ転送時間を調べた。これがほぼ 4.2msecであり表1の計算時間に比べれば無視し得る範囲にある。

8.4.2 負荷配分

異なるSCモジュールは前段の異なるSCモジュール集合からデータを受け取る。前段のSCモジュールがゼロでない出力を出すかどうかは入力パターンに依存する。従って、一つのSCモジュールが受け取るデータ量はダイナミックに変動する。これはSCモジュール内に負荷の不均衡を生む。特に第3、第4といった奥のステージでの不均衡は大きい。一つの解決策としては次のような方法が考えられる。一つのステージにおけるSCモジュールはその前段のモジュールの類似性に従ってクラス分けできる。異なるクラスに属するSCモジュールは、それぞれの前段モジュールについて重複しない。ノードのクラスタがこれらのSCモジュールを共同して処理するなら、各ノードの負荷は入力パターンに依らずに平均化する。もう一つの解決策は、flood fill algorithm⁽⁸⁾のようなダイナミックなノード割り当てである。この方法においては、「管理者ノード」が「労働者ノード」に逐一データパケットを配分する。

8.5 おわりに

以上、ネオコグニトロンをNCUBEにインプリメントした結果について述べた。現在以下のような課題を検討中である。

- (1) 同時に処理する文字数を増やす
- (2) テレビカメラから直接入力した文字を読み取る
- (3) 認識対象をアルファベットに拡張する

8.6 文献

- (1) 三宅誠,伊藤崇之,福島邦彦:神経回路モデルの並列処理—NCUBEによるネオコグニトロンの高速実行—,昭和63秋季信学全大講演論文集,p.D-1-9(1988)
- (2) 伊藤崇之,福島邦彦,三宅誠:並列計算機によるネオコグニロン型数字認識システムの実現,昭和63秋季信学全大講演論文集,p.D-1-10(1988)
- (3) T.Ito,K.Fukushima,S.Miyake:Pattern Recognition and Parallel Processing:NEOCOGNITRON on NCUBE, 2nd International Symposium on Bioelectronic and Molecular Electronic Devices, R&D Association for Future Electron Devices, Dec.12-14, Fujiyoshida(1988)
- (4) K.Fukushima:Neocognitron: A self-organizing multilayered neural network, Biol. Cybern., 36, No.4, pp.193-202(1980)
- (5) W.D.Hillis:The Connection Machine,MIT Press(1985)
- (6) K.Fukushima: Neocognitron: A Hierarchical Neural Network Capable of Visual Pattern Recognition, Neural Networks, 1, No.1, pp.119-130(1988)
- (7) H.J.Siegel, R.J.McMillen:The Multistage Cube: A Versatile Interconnection Network, IEEE Computer, 14, No.12, pp.65-76(1981)
- (8) Parallel Fortran User Guide, 3L Ltd.(1988)

9. 3層逆伝播学習ネットワークによる画像データ圧縮

9.1 はじめに

近年コンピュータ処理速度の向上を背景に、変換規則を自動的にサンプルデータから抽出する学習能力と、複雑な処理を高速に実現する同時並列処理性、の2つの特徴に着目したニューラルネットの工学的応用研究が進展しつつある。

ニューラルネット応用の一つに視聴覚パターン処理がある。これらの応用では、より小さな次元の特徴空間へ変換する機能（情報圧縮性）と、学習データから規則性を抽出して未知の類似データに対して学習データと同様の性能を得る機能（汎化性）が重要である⁽¹⁾。

一方、画像圧縮技術は、再生画像の品質を維持しつつ信号の持つ冗長度を抑圧してデータを圧縮することを基本とする。これまで主に、輝度、色相等の信号レベルや画素の2次元空間分布に関する統計的冗長度を抑圧する符号化が検討されている。さらに人間の視覚・心理特性に基づく視覚的冗長度⁽²⁾や、送受信で共有する対象の構造モデル等の知識的冗長度⁽³⁾を抑圧する方式も研究されてきている⁽⁴⁾。

本章では、再生画像の2乗平均誤差を最小にするという拘束条件のみを与えたニューラルネットを用い画像のデータ圧縮を行い、ニューラルネットの情報圧縮と汎化能力について評価する。また、学習によって自動的に獲得した圧縮・再生規則の性質や、ネットワーク内でのデータ表現方法を検討し、新たな画像の符号化方式を探る。この種の研究には、逆伝搬学習則(Back-propagation)を用い単一ネットを画像に適用した検討⁽⁵⁾や、局所並列的に適用した検討⁽⁶⁾等の報告がある⁽⁷⁾、⁽⁸⁾。

今回は、3層の非巡回形ニューラルネットと逆伝搬学習則を用い、静止画像を対象としたデータ圧縮について報告する。しかしながら現在、ニューラルネットを構成するパラメータ、圧縮特性、記憶容量、学習の収束性、実現する写像関数の精度、汎化特性等をシステムティックに決定できるような設計原理は確立していない。このため、計算機シミュレーションによる実験的検討を行った。主な内容は、まず、画像データをニューラルネットで高速処理するプログラムの、超立体アーキテクチャを用いた汎用並列計算機(NCUBE)上での実現方法、得られた性能について述べる。次に、これを用いた画像圧縮実験結果として、学習画像数と再生した学習画像のSN比特性、未学習画像に対するSN比特性、中間層出力の量子化に対するSN比特性、重みの初期値に対するSN比特性について述べる。

9.2 ニューラルネットによる画像圧縮モデル

9.2.1 システム・モデル

人工的ニューラルネットを用いて、画像信号の圧縮・再生を行う符号化方式を、ここではニューロ・コーデック(Neuro-CODEC)と呼ぶこととする。そのシステム・モデルの概略を図1に示す。ニューラルネットには、入力画像をより低次元のデータに変換して圧縮機能を実現させることが要求されるので、入出力層のユニット数に比して中間層のユニット数は小さくする。

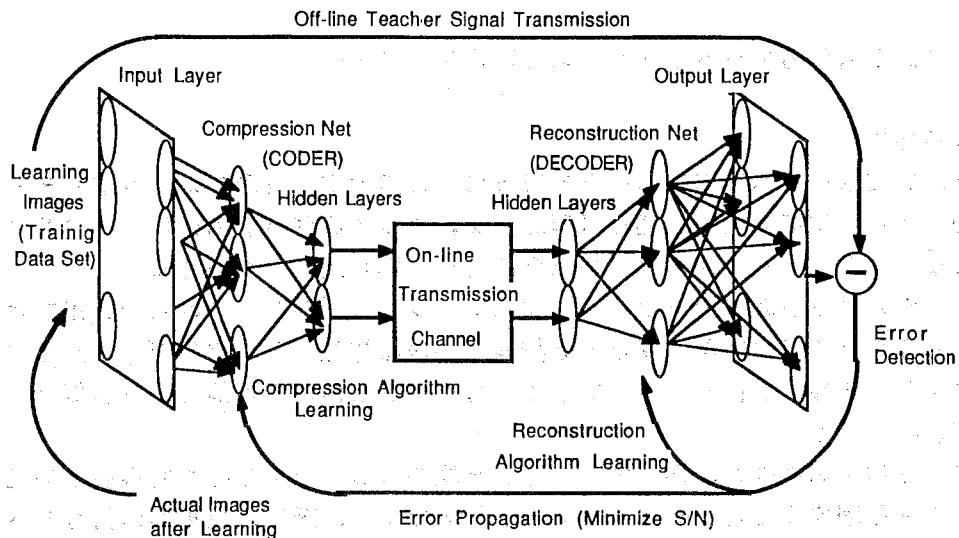


図1. Neuro-CODEC のシステム・モデル概要

学習過程では、学習すべき画像データを入力層に与え、出力層には教師信号として入力層と同一の信号を与える。ニューラルネットは、圧縮ネット(Compression Net)と再生ネット(Reconstruction Net)からなる。圧縮ネットは、サンプルデータから画像信号の持つ冗長度を抑圧する変換規則を学習する。再生ネットは、圧縮されたデータから原画像を再生する変換規則を学習する。

なお、通信過程については、ニューラルネットを中間層を中心にして、圧縮ネットと再生ネットに分離して回線の送受信符号器(CODER)、復号器(DECODER)として通信路を構成することになる。

9.2.2 ニューラルネットのモデル

実験に用いたニューラルネットの構成は、図2に示す3層の非巡回型ネットワークであり、学習則は誤差逆伝搬法を用いた⁽⁹⁾。

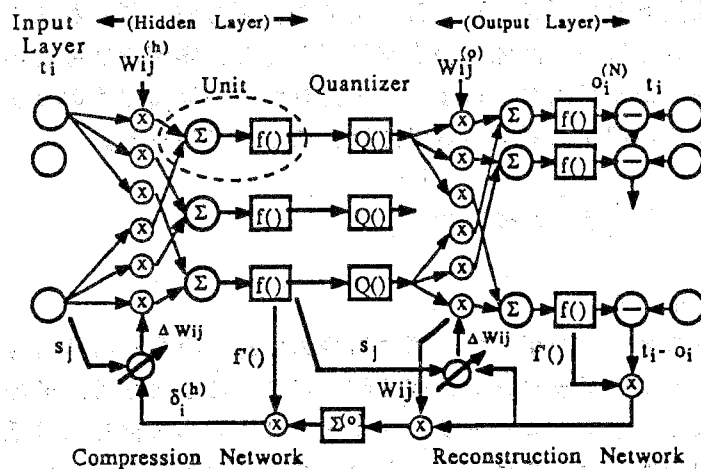


図2. ニューラルネットの構成

(1) ネットワーク構成

第 n 層 i 番目のユニットの入力 $net_i^{(n)}$ は、重み係数 $w_{ij}^{(n)}$ と $(n-1)$ 層の j 番目のユニット出力 $s_j^{(n-1)}$ を用いて、次式で示す線形結合演算により得られる。

$$net_i^{(n)} = \sum_j w_{ij}^{(n)} s_j^{(n-1)}, \quad n \geq 2 \quad (1)$$

ここで、 $net_i^{(n)}$ のバイアス値は、 $s_j w_{ij}$ ($s_j=1$) として(1)式に組み込む。第 n 層の活性化関数としては次式のシグモイド関数を用いた。

$$o_i^{(n)} = f(net_i^{(n)}) = 1 / [(1 + \exp(-net_i^{(n)}))] \quad (2)$$

再生ネット出力 $o_i^{(N)}$ と教師 t_i との誤差は次式で示す2乗誤差関数を採用した。

$$ERROR = (1/2) \sum_i (t_i - o_i^{(N)})^2 \quad (3)$$

(2) 学習則

第 (n) 層での i 番目のユニット入力値 net_i に対する誤差 $ERROR$ の変化分 $\delta_i^{(n)}$ は次式となる。

$$\delta_i^{(n)} = d \text{ ERROR} / d net_i^{(n)} = [s_j^{(n)} (1 - s_j^{(n)})] \cdot \sum_i \delta_i^{(n+1)} w_{ij}^{(n+1)} \quad (4)$$

学習時に於ける重みの変化分は以下によって修正する。

$$\Delta w_{ij}^{(n)} = \alpha \delta_i^{(n)} * s_j^{(n-1)} \quad (5)$$

次に α, ϵ を重み修正利得係数、 p をパターン数とすると重みの修正は以下となる。

$$\Delta w_{ij}^{(n), p} = \alpha \Delta w_{ij}^{(n), p} + (1 - \alpha) \delta_i^{(n)} * s_j^{(n-1)} \quad (6)$$

$$w_{ij}^{(n)} = w_{ij}^{(n)} + \epsilon \Delta w_{ij}^{(n), p} \quad (7)$$

ここでは、 $\epsilon, p = 1$ とし1学習毎に $\Delta w_{ij}^{(n)}, w_{ij}^{(n)}$ を変更し、 $\Delta w_{ij}^{(n)}$ は全学習での累積値を用いた。

(3) 画像品質評価尺度

画像品質の評価尺度は、次式で示す規格化2乗平均誤差を用いた。

$$SNR = 10 \log_{10} [E\langle t_i^2 \rangle / E\langle (t_i - o_i)^2 \rangle] \quad (8)$$

ここで、 $E\langle \rangle$ は、画面全体での平均を示す。

9.2.3 並列計算機によるインプリメント

ニューラルネットは、並列・多層構造と最急降下法によるコネクシオン重みの適応処理を基本とするため演算量が膨大である。特に、画像を対象とする場合、ユニット数及びコネクシオン数の増加と長大データの入出力時間が大きくなり、長時間の計算を要する。このためシミュレーションには、超立体アーキテクチャの汎用並列計算機（NCUBE）を用いた⁽¹⁰⁾。この種の並列計算機は、本来ノードプロセッサ間の通信が少ないことを前提に高性能が得られるように設計されている。

そこで、画像が離れた画素間での相関が少ない性質を用い、図3に示す局所並列形(Local-Parallel NN)を採用した。これは、画像をブロック分割し、ブロックごとに独立な小ニューラルネットを適用し、同時並列処理する。このため、小ニューラルネットは、顔画像であれば、目、鼻、髪、背景等の各部位の特徴に対応する圧縮・再生アルゴリズムを学習することになる。また、小ニューラルネット間での通信をなくすため、ブロックが重複した部分は複数の小ニューラルネットで処理する形式とした。

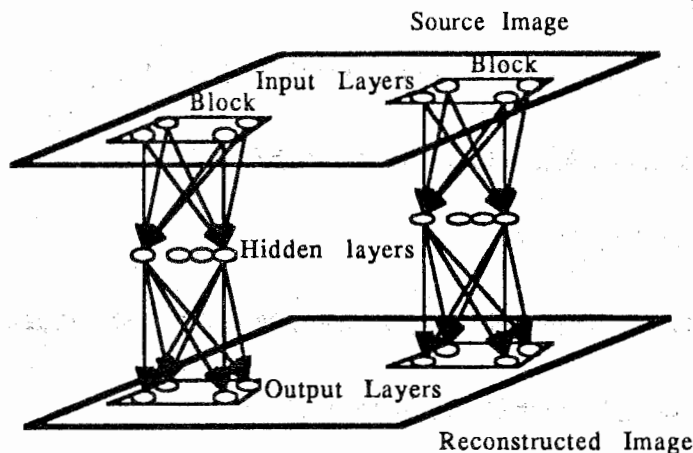


図3. 画像圧縮のためのニューラルネット型情報処理

次に、このニューラルネットを実現する時、画像を蓄積するホストプロセッサとノードプロセッサ間でのデータ転送時間をなくし、ノードプロセッサの実効処理効率を向上させるため以下のノードプロセッサ機能配分を行った。図4に示すようにノードプロセッサを、①フォワード演算及びバックプロパゲイション演算を行うカーネル処理と、②プロセス制御、同期及び学習データの蓄積を行うフロント・エンド処理、とに分割するアーキテクチャとした。この方式においては、ニューラルネット処理時にノードプロセッサの局所メモリ（最大512kB）に画像データを常駐させ、学習用の画像ブロックデータをカーネル処理ノードに分配転送する。この結果、ノードプロセッサの稼働率約80%を得た。

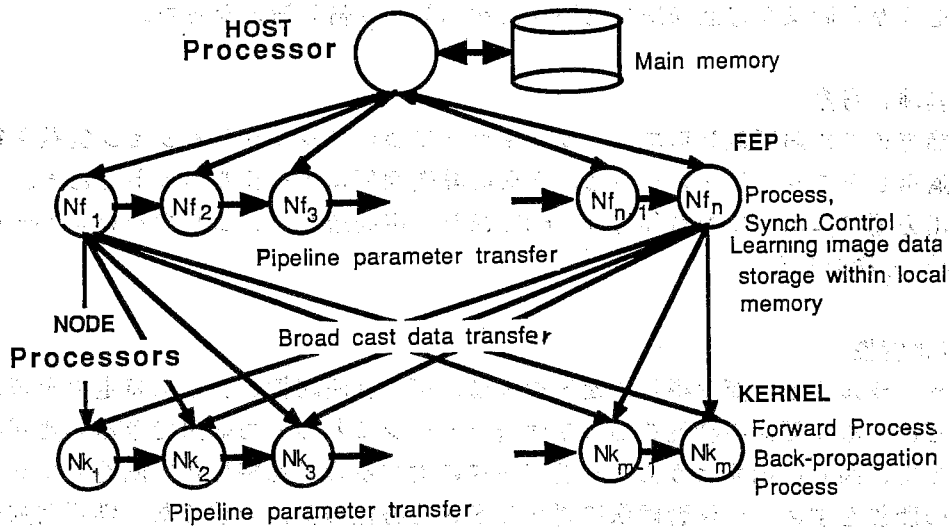


図4. ノードプロセッサへの機能割り当て

本方式の処理能力を図5に示す。図5は、3層の小ニューラルネット（ニューロン数 64×3 、コネクション数 $64 \times 64 \times 2 = 8192$ ）の処理を単一のノードプロセッサで実行する場合の処理時間を示している。信号の表現精度は、入出力画像が 128 画素 $\times 128$ 画素 $\times 8$ bits、ニューラルネット内部演算: 4 bytes 実数である。

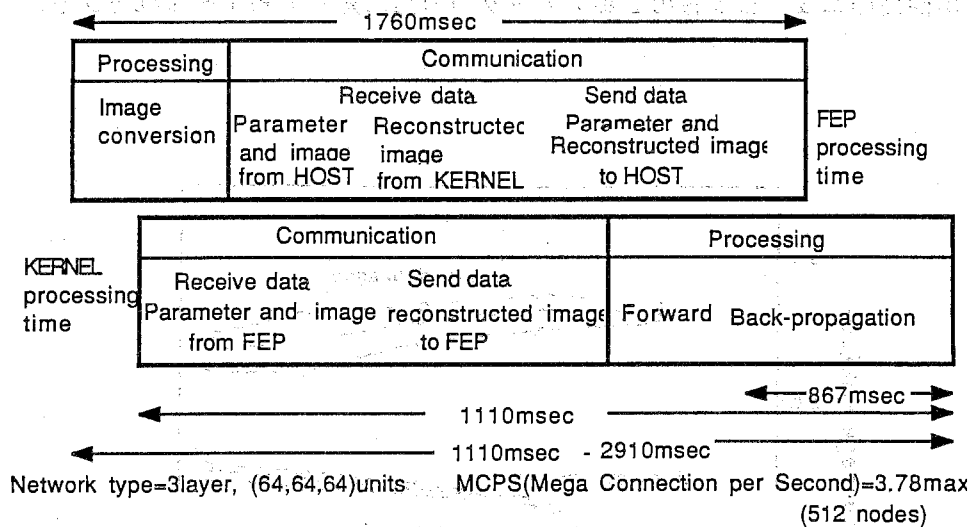


図5. 一つのノードプロセッサにおける処理速度

処理能力については、512個のノードプロセッサを用いて、これと同数の小ニューラルネットにブロック分割した単一画像を同時並列で学習させる場合、3.78 MCPS(Mega Connection Per Second)の性能を得ている(11)。

一方、学習毎に画像を更新する場合、フロントエンド・プロセッサからカーネル・プロセッサに画像ブロックデータが毎回転送される。この時の処理能力は、512個で1.44 MCPSである。この結果、1枚の画像を約1秒で学習可能となり、ホス

トメモリから画像を毎回転送した場合に比べ約20倍高速となった。

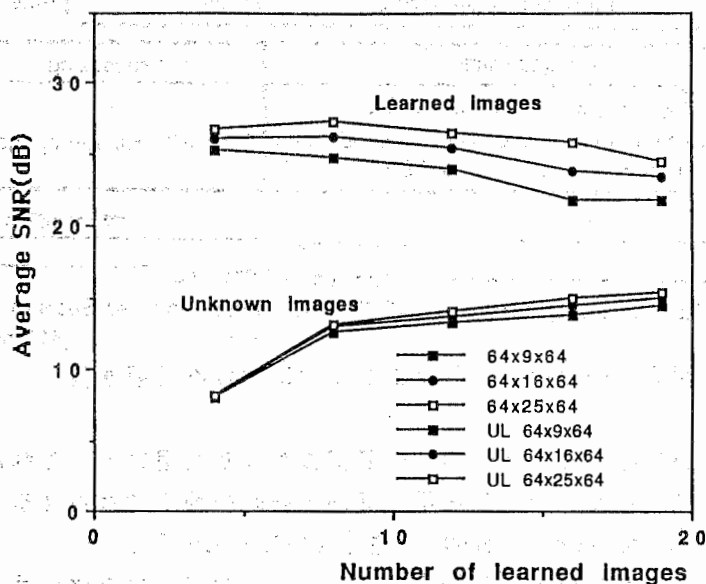
9.3 実験と考察

2節で述べた局所並列形ニューラルネットについてシミュレーション実験を行った。適用したニューラルネットは、3層の非巡回型ニューラルネットである。入力層と出力層のニューロン数(N_i)は、64であり、中間層を N_i より少ないニューロン数(N_h) で構成した。

(1) 汎化特性

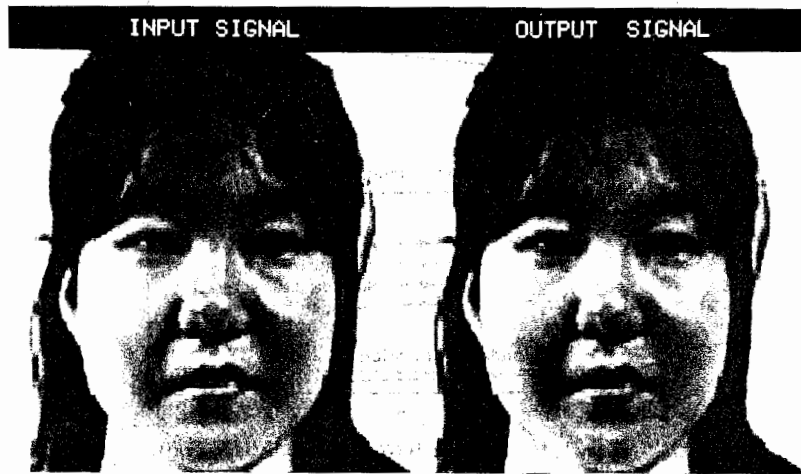
ニューラルネットが獲得した変換規則は、データ圧縮性ととも特定の学習パターンだけでなく類似の特性をもつ未学習パターンに対しても適用できる必要がある。

そこで、図6に示すように、学習データに対するSN比特性と中間層のニューロン数の関係を求め、未学習画像に対するSN比特性の傾向を求めた。学習回数は、各学習画像に対して500回であり（全体の学習回数は学習画像数5-19枚に対し2500-9500回）、学習後の再生画像のSN比は、学習画像全体での平均値である。また、図7に学習画像と学習後の再生画像例を示す。学習画像数の増加に伴い、再生画像のSN比は減少するが、平均SN比は、中間層のニューロン数(32,42,52)が多いほどSN比は高い。これは、ニューラルネットがパターン情報を重みベクトルの形で圧縮して記憶する能力に対応するものと考えられる。また、SN比の増加は、中間層のニューロン数の2乗のオーダーの増加に対して約1~2dBである。

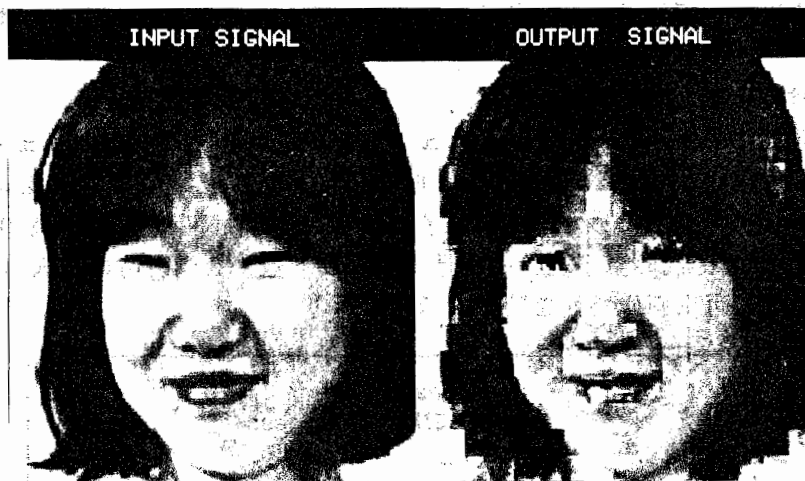


学習回数:500回/1画像 量子化ビット:8 bits

図6. ネットワークの汎化能力



学習画像とその再生例 SNR=23.74db



未学習画像とその再生例 SNR=15.55db

学習回数:500/image, 量子化ビット:8bits, 回路:64×16×64(3層), 画像数:19

図7. 再生画像の例

一方、未学習画像に対する平均SN比は学習画像数の増加に応じて向上する。図6に示した未学習画像に対するSN比は、未学習画像3画面についての平均値であり、図7の下段に未学習画像とその再生画像例を示す。図7から分かるように再生画像にはブロック歪が現れ、SN比劣化要因となっている。未学習画像の平均SN比は、中間層のニューロン数16の時、学習画像数が4で8.14dB、19で14.92dBであった。学習画像19枚を用い、学習回数を9500から38000まで行った場合の再生画像の特性を図8に示す。学習画像に対しては上記の結果と異なり、SN比が約3.7dB向上するが、未学習画像に対するSN比は若干劣化する。これは、学習画像の特性にチューニングされ過ぎる過学習効果のためであると考えられる。

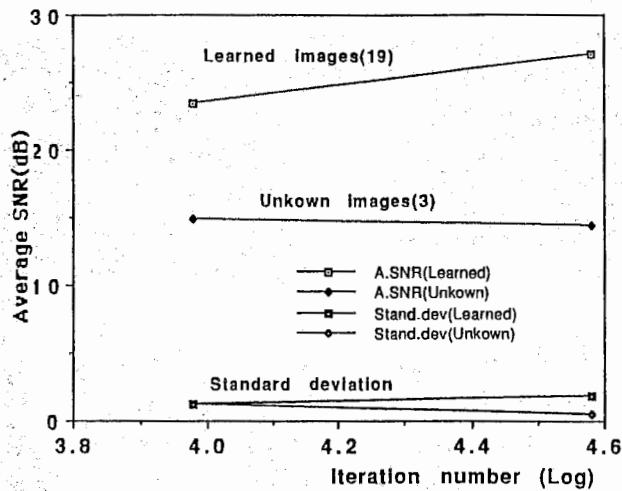


図8. 学習回数に対する画像品質 初期値:ランダム, 回路:64×16×64, $\alpha=0.9$
汎化性能を上げるには、直交変換等で次元を減らす必要がある。また、ブロック歪軽減のためには、画像ブロックをオーバーラップさせる等の処理が必要である。

(2)量子化特性

データ圧縮では、入力層と出力層のニューロン数 (N_i) と中間層のニューロン数 (N_h) の比とともに、中間層での量子化精度が問題となる。そこで、図2に示すように中間層の出力全てに対して、一様量子化器(シグモイド関数の出力値 0.0-1.0 を 2^{-n} ステップで等間隔量子化、 n は量子化ビット数)を学習時に挿入した。量子化精度に対する再生画像特性を図9に示す。

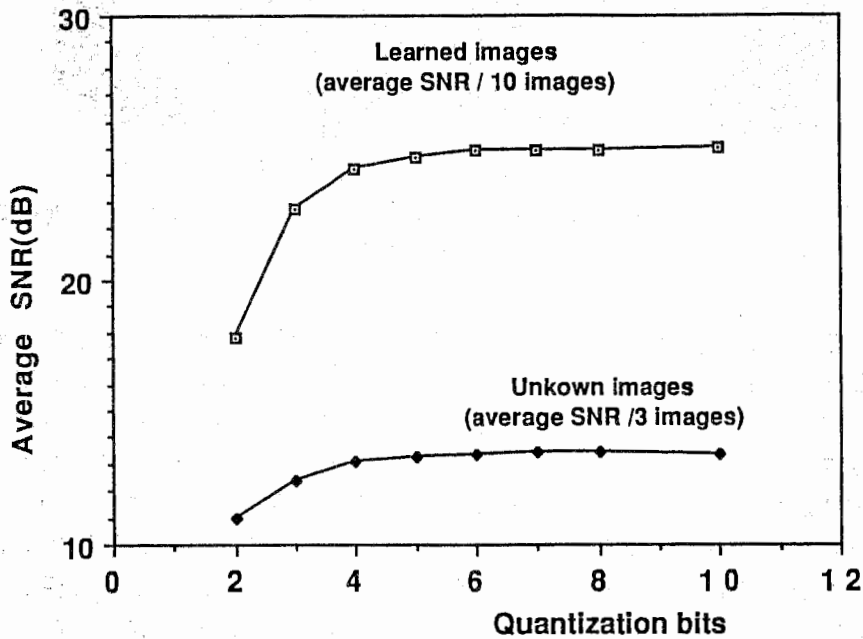


図9. 量子化精度に対する画像品質 学習回数:400回/image, 回路:64×16×64

学習時と学習後に中間層出力を n ビット量子化する場合を比較すると、学習時に挿入した場合の方が良い結果を得る。 $n=3$ とすると、学習時量子化では約 2dB 劣化であるのに対し、学習後では約 6dB 劣化であった。これは、学習により 2 乗平均誤差を最小にするよう量子化入力信号分布を変化させているためと考えられる。

本稿では、シグモイド関数が圧伸特性を持つことから、中間層の出力全てに対して同一ビットの線形量子化を用いたが、非線形量子化及び各ニューロンの出力に対して最小量子化ビット数となるよう学習させる検討を行う。

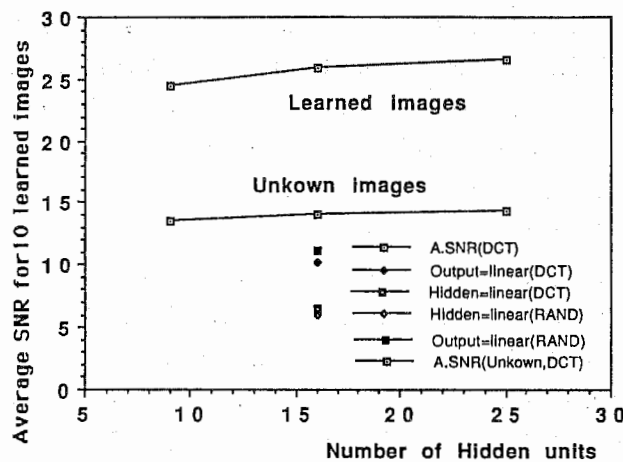
(3) 重み初期値

従来、ニューラルネットの接続の重み初期値は、一様乱数が用いられていた。しかし、図 2 の回路を見ると、中間層の積和演算では直交変換を並列演算し、出力層で逆変換しているものとも見なせる。ただし、出力値に非線形のシグモイド関数が挿入されており、学習によって重み係数を最急降下法で適応制御しているので、初期値を適当に選べば非線形が多層適応型直交変換符号化と考えても良い。あるいは、初期値として画素間の自己相関係数を p_{nm} を選べば、並列演算タイプの多層適応型線形予測符号化とも見なせる。そこで、離散コサイン変換核 (DCT) を重みの初期値とした場合について検討した。重みの初期値として、

$$w_{k,l} = \cos[(2k+1)m\pi/2M]\cos[(2l+1)n\pi/2M]$$

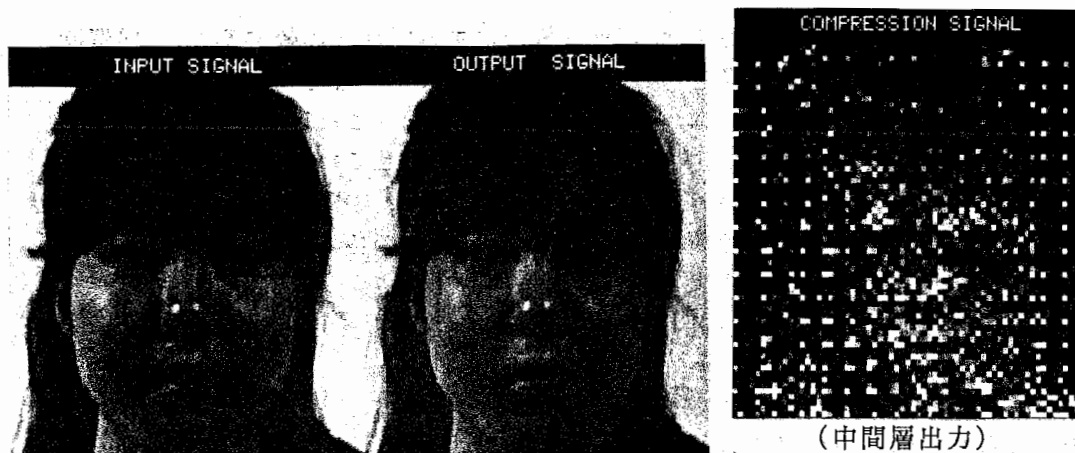
を採用する。ここで、 k, l は入出力層ニューロン、 m, n は中間層のニューロンを示す。

図 10 に、DCT 変換核を重み初期値として学習した場合の特性を示す。低次の係数のみを用いる直交変換符号化では、汎化特性は殆ど問題にならない。しかし、図 6 と比較すると、ニューラルネットが収束状態での学習、未学習画像に対する SN 比は、一様乱数の場合とほぼ同一であった。中間層でのニューロンの発火状態を図 11 に示す。同図に示すように、初期状態からは学習によって変形するが、 8×8 ブロックの左下が直流分、右上が周波数面での高次スペクトル成分に対応する。このため、初期値を DCT 変換核とすることでスペクトル成分での学習が可能となる。



学習回数: 400回/image, 回路: $64 \times 16 \times 64$

図 10. 初期値として DCT 変換核を採用した場合の再生画像品質



学習回数:400回/image, 回路:64×16×64, 学習画像数:10

図1.1. 初期値としてDCT変換核を採用した場合の学習済み画像の再生例

9.4 まとめ

- ① バックプロパゲーション学習則を用いた高速な画像処理用ニューラルネットを並列計算機NCUBE上に実現し、約1秒で画像を学習できる性能を得た。
- ② 圧縮特性として、学習画像数、中間層出力の量子化、重み初期値、それぞれに対するSN比特性をシミュレーションにより求めた。

9.5 文献

- (1) 上坂:ニューラルネットの基礎論,TAIS,No.2,pp.2-30(1988.06)
- (2) M.Kunt,A.Ikonomopoulos,M.Kocher:Second-Generation Image-Coding Techniques,Proc. IEEE,Vol.73,No.4,April(1985)
- (3) 原島博:知的画像符号化と知的通信,TV誌,Vol.41,No.7,pp.519-525(1988)
- (4) 田崎,山田:動画像符号化の体系化とその応用,信学誌,Vol.71,No.7,pp.663-668(1988)
- (5) G.W.Cottrell,P.Munro,D.Zipser:Image compression by Back Propagation: An Example of Extensional Programming, ICS Report 8702, February(1987)
- (6) 曾根原,川人,三宅,中根:ニューラルネットによる画像データ圧縮(Neuro-CODEC)の検討,信学技報,Vol.88,No.195(1988)
- (7) J.G.Daugman,D.M.Kamnen:Image Statistics, Gases, & Visual Neural Primitives, ICNN June, IV-163(1987)
- (8) L.D.Jackel,R.E.Howard,et al:Building a Hierarchy with Neural network :an Example Image Vector Quantization,Vol.26,No.23,APPLIED OPTICS(1987)
- (9) D.E.Rumelhart,G.E.Hinton,R.J.Williams:Learning representations by back-propagation errors, NATURE, Vol.323, 9 October (1986)
- (10) 三宅他:神経計算論と並列信号処理,『インテリジェント信号処理』,pp.1-10(1987,12)
- (11) G.Blelloch,C.R.Rosenberg:Network Learning on the Connection Machine", IJCAI,Knowledge Acquisition,pp.323-324(1987)

10. 前注意過程における図形の特徴点抽出モデルと画像復元

パターン認識の能力は対象の表現方法に大きく依存している。人間は高度なパターン表現機能を持っており、人間のパターン表現方法、記憶情報のアルゴリズム、パターンの学習アルゴリズム等を知ることが、工学的にもきわめて有意義である。これまで認知心理学を初めとする心理学の領域で人間のパターン認識のさまざまな特性がしらべられてきた。しかしながら、ほとんどの心理学的研究は情報処理過程の時間特性や容量をしらべたものであり、情報の内部表現すなわち脳内表現について検討したものはきわめて少ない。本研究はこれまで明らかにされてきた心理学的ならびに生理学的知見を踏まえ、パターンの脳内表現とりわけ2次元曲線(図形)の記憶表現について考察を行なう。

10.1 本論文の構成

本稿では、まず曲線の表現に関する計算論的考察を行なった後、曲線の「特徴点」を定義し、曲線の特徴点を並列に検出するアルゴリズムを提案する。次に曲線の構造記述という立場からアルゴリズムを改良し、パターンの大局的記述に関してシミュレーションを行なう。最後に得られた特徴点から元の曲線の復元を試みる。

10.2 計算論的考察

数学では、曲線の方程式を与えられたとき、その大域的形状を推測することを曲線の追跡と呼ぶ。この場合、大域的形状とは

対称性

特異点の位置、その付近の形状

曲線の凹凸、変曲点

などが挙げられる。人間のパターン認識ではもちろん曲線の方程式が与えられるわけではないが以上のような特徴は重要であり、何らかの方法で検出しているにちがいない。

2次元曲線の構造記述はパターン認識においてきわめて重要な問題である。人間の視覚研究においても、まず網膜像の強度変化が検出されその局所的な幾何学的構造が表現されると考えられている。この表現は原始スケッチ(primal sketch)と呼ばれている⁽¹⁾。原始スケッチの重要な機能は、画像の強度変化の検出と局所的な幾何学的構造の表現にある。原始スケッチは、素原始スケッチ(raw primal sketch)がと完全原始スケッチ(full primal sketch)に分けられている。素原始スケッチでは、縁edge、棒bar、端点termination、小塊blobといった表現素を用いて画像を表現する。

一方これまで曲線の特徴づける方法としてさまざまな方位を持つ線検出器によって曲線の各位置における接線方向を計算する試みがなされてきた⁽²⁾。また、複数の長さの線検出器から曲線の端点を求める方法も考えられてきた^{(3),(4)}。2次元

曲線の表現に関する研究はこれまで数多くなされてきたが、まず曲線の特徴点を検出し、それらの関係を記述することが基本的な方法だと思われる(5)、(6)。

そこでパターン認識に重要な役割をはたしていると思われる曲線の「特徴点」を定義し、それらを並列に検出するアルゴリズムを提案する。2次元曲線の重要な特徴として曲率の大きい部分 (high curvature point)、曲線の交点 (cross point)、曲線の端点 (termination)、間隙 (gap) などが考えられる。これらは、曲線の特徴づける重要な位置であり特徴点 (臨界点) と呼ばれている。曲率の大きい部分は心理学的にも工学的にも重要であることが認められている。例えばAttneave(7) は心理学的実験から、曲率の大きい位置は情報量が高いことを示唆した。また HoffmanとRichards(8) は図形の部分分割法として、曲率の大きい凹の部分で分割するのが自然であることを見い出している。一方心理学的研究から、人間が特定の位置に注意を払う前に (前注意過程において) 並列に処理できる特徴があることが示唆されている。Juleszは並列処理可能な特徴をテクストン (texton) と名づけた(9)。テクストンの構成要素は線分の有無・長さ・幅・方位、線分の端点、線分の交点、小塊の有無・長さ・幅・方位などがある。興味深いのはこれらの構成要素の多くが上記の特徴点の特徴と一致している点である。以上のことから曲線における特徴点は曲線の特徴づける上で重要であるばかりでなく、前注意過程において並列に処理されるらしい (図1)。さらに、これらの特徴点付近の特徴だけでもとの図形情報はかなり精度よく復元できることが予想される。

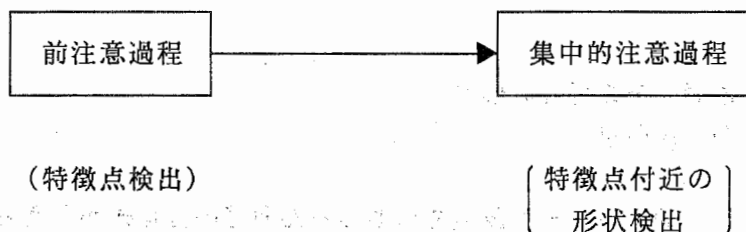


図1. 形状の階層的処理

10.3 曲線の臨界点検出に関する並列アルゴリズム

ここでの基本的な仮定は曲線の詳細な特徴を線検出器などを用いずに方向選択性の低い微小なチャンネルによって検出されたエッジ情報によって知覚されるということである。すなわち以下に示すような並列アルゴリズムによって臨界点をまず検出し、次に注意過程を通して臨界点付近の詳細な形状をこの微小なチャンネルからのエッジ情報に基づいて知覚すると考えるのである。

(1) 計算論的ならびに生理学的知見

雑音が含まれる画像の中からできるだけ正確に線分を検出するための画像フィルタが研究されてきた。Canny(10) は、変分法を用いて最適フィルタの研究を行った。

Canny フィルタは、基本的には2次元ガウス関数の一階微分の形をしている。

一方、高等動物の脳視覚領には、方向選択性の低いきわめて小さな受容野を持つ細胞が高密度で存在していると同時に方向選択性の高い複数の大きさの受容野を持つ細胞も存在することが生理学的研究から明らかにされている。さらに人間の視覚領には複数の大きさの線検出器が存在することが知られている。

(2) 並列アルゴリズム

我々はここでまず曲線データを方向選択性の低い微小なチャンネルと方向選択性の高い線検出器（以後単に線検出器と呼ぶ）によって、同時並列的に処理すると仮定する。次に同じ位置の情報を処理する線検出器から方向選択性の低い微小なチャンネルに対して抑制性の出力がかかると仮定する。このように仮定すると曲線の直線性の高い部分では線検出器の出力は高くなり方向選択性の低い微小なチャンネルの出力は抑えられる。一方直線性の低い部分では線検出器の出力が低く、微小なチャンネルの出力にはあまり影響を与えないであろう。その結果、直線性の低い部分、すなわち臨界点付近においてのみ微小なチャンネルの出力が得られることが予想される。図2は処理の流れ図を示している。まずステップ1ではエッジが抽出される。この段階での出力は2次元曲線の詳細な形状情報を与えている。次にステップ2ではこの出力に方向性を持つ線検出器から抑制がかかり臨界点がマークされる。

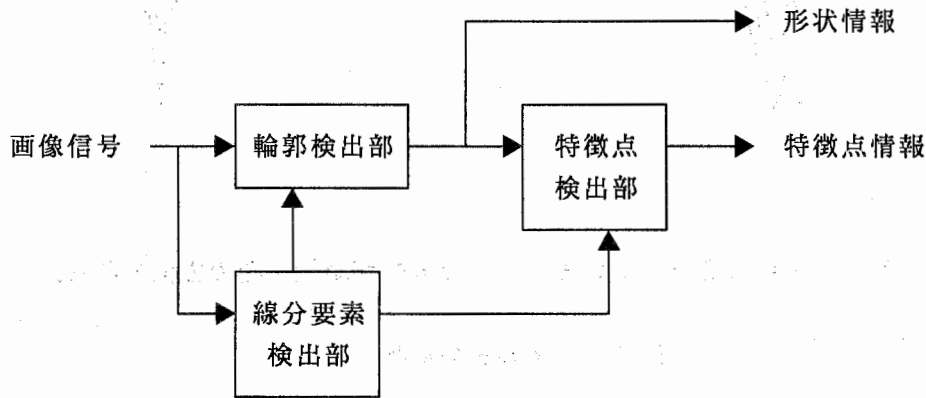
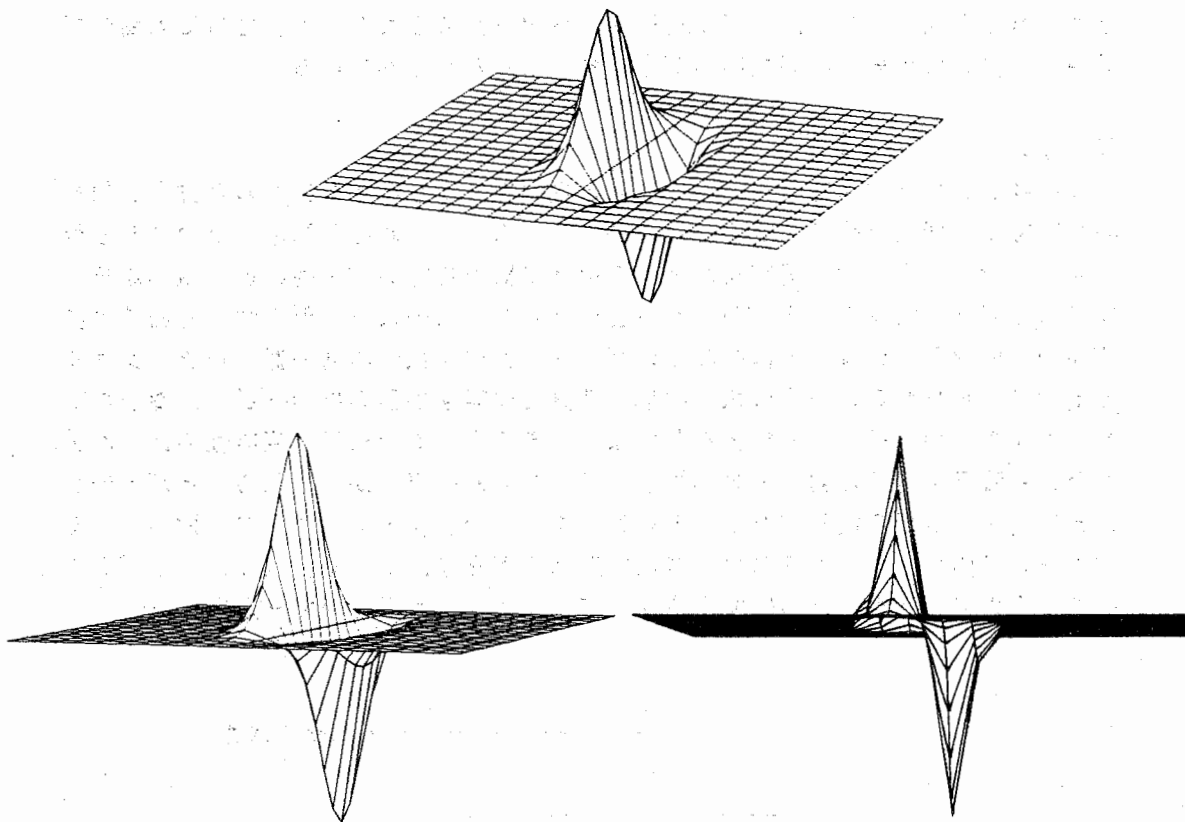


図2. 処理の流れ図

次に各チャンネルのフィルタ特性について説明する。ここでは方向選択性の低い微小チャンネルと線検出器が同一の関数で決まるフィルタ特性を持つと仮定した。ここでは、軸に垂直な方向の断面がガウス関数の一階微分の形をし、軸に平行な断面がガウス関数になっているフィルタを用いた（図3）。すなわち、フィルタの空間特性 $f(x, y)$ は、

$$f(x, y) = x \exp(-x^2/2\sigma_x^2 - y^2/2\sigma_y^2) / (2\pi \sigma_x^3 \sigma_y)$$



図は $\sigma_y = 3 \cdot \sigma_x$ の場合を示している。 σ_y の方位が線検出器の方位である。

図3. フィルタの形状

これらのチャンネルのフィルタ特性は、ガウス関数の一階微分の形をしているので、 σ_x の値を変えることによって、画像に対してさまざまな度合で平滑化した後、微分操作を施すことにより強度変化の大きい部分を検出することができる。ここで方向選択性の低い微小なチャンネルの場合、

$$\sigma_x = \sigma_y$$

とし、線検出器の場合、

$$\sigma_y = n \sigma_x \quad (n \text{ は整数})$$

とした。図3にフィルタの空間特性 $f(x,y)$ の一例を示す。 n を変化させると線検出器のフィルタの長軸方向の長さが変わり直線性の判定基準が変化し、検出される曲率の範囲が変化する。すなわち n を大きくすることによって検出される曲率の範囲が広がる。

次に線検出器の閾値は各フィルタの最大出力の $1/2$ とした。このように閾値を設定すると直線に近い部分だけが出力されその結果方向選択性の低いチャンネルの出力は曲線の臨界点付近だけが残ることが予想される。

具体的なアルゴリズムは以下の通りである。画像データが与えられると二種類のチャンネルによって同時に処理が始まる。微小チャンネルのマスクサイズは 7×7 (フィルタの形状を 7×7 で表現)、サンプリングは 3×3 おきに行なった。 $\sigma_x = \sigma_y = 2.0$ で方位は 10° おきに36種類のフィルタを用意した。次に、閾値を2.5として出力が閾を越えた点をエッジとして表示した。このようにして抽出されたエッジ部分について

$$\sigma_x = 2.0, \sigma_y = 6.0$$

$$\sigma_x = 2.0, \sigma_y = 7.0$$

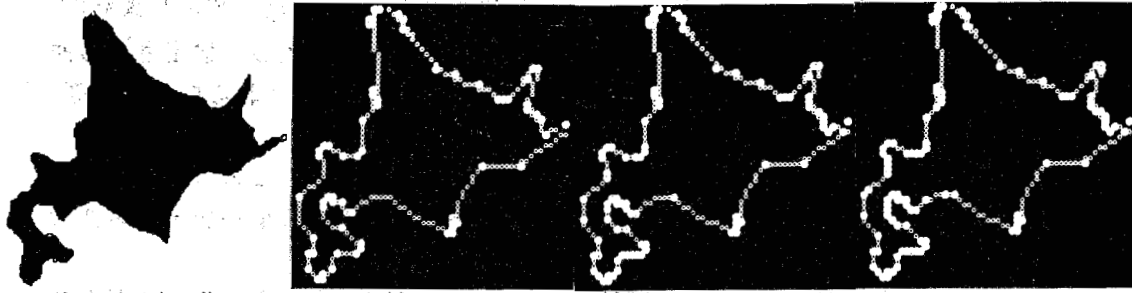
$$\sigma_x = 2.0, \sigma_y = 8.0$$

の3種類の線検出器によって再びコンボリューションを行なった。

線検出器の最大出力は2値(0,-1)パターンの場合、 $\sqrt{2\pi}\sigma_y$ となる。したがって上記3種類のフィルタに対して、最大出力の $1/2$ すなわち、7.56、8.82、10.08を閾値とした。ここで閾値を越えなかった部分で方向性の低いフィルタの出力が得られている部分を特徴点としてマークした。

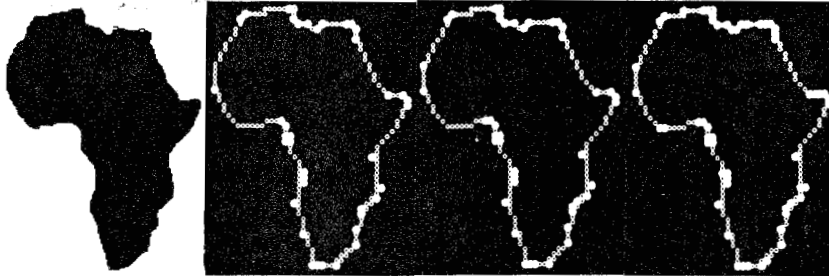
10.4 並列計算機NCUBEによるシミュレーション (曲線の臨界点検出)

NCUBEは512個のプロセッサがハイパーキューブ上に接続されている並列計算機であり、これを用いてシミュレーションを行った。ディスプレイは 256×256 画素からなりディジタイザを用いて図形入力を行った。図形部分の画像データは-1であり背景は0である。 256×256 画素に対して 3×3 のサンプリング間隔で処理を行なった。検出器の方位は36種類である。上述のようにノード(プロセッサ)数は512個なので、画像の1行(256画素)を6個のノードに割り付けて処理を実行した(各ノードは6方位を処理する)。計算時間は約2秒であった。入力図形として北海道とアフリカ大陸を用いた結果を図4および図5に示す。各 σ_y に対して抽出されたエッジ(灰色)と検出された臨界点(白丸)が示されている。つぎに線図形を入力したときの結果を図6に示す。曲率の高い点以外にも、端点、交点等の特徴点がうまく抽出されている。また σ_y を大きくすることによって曲率の小さい部分まで検出されていることがわかる。このことは、量子化ノイズがあり鮮明ではないものの図7からも明らかである。



左から原画像、 $\sigma_y=6.0$ の結果、 $\sigma_y=7.0$ の結果、 $\sigma_y=8.0$ の結果。それぞれエッジの数219に対して、64、88、100個の特徴点が抽出された。

図4. 北海道を入力図形とした場合の特徴点



左から原画像、 $\sigma_y=6.0$ の結果、 $\sigma_y=7.0$ の結果、 $\sigma_y=8.0$ の結果。それぞれエッジの数154に対して、40、44、57個の特徴点が抽出された。

図5. アフリカ大陸を入力図形とした場合の特徴点

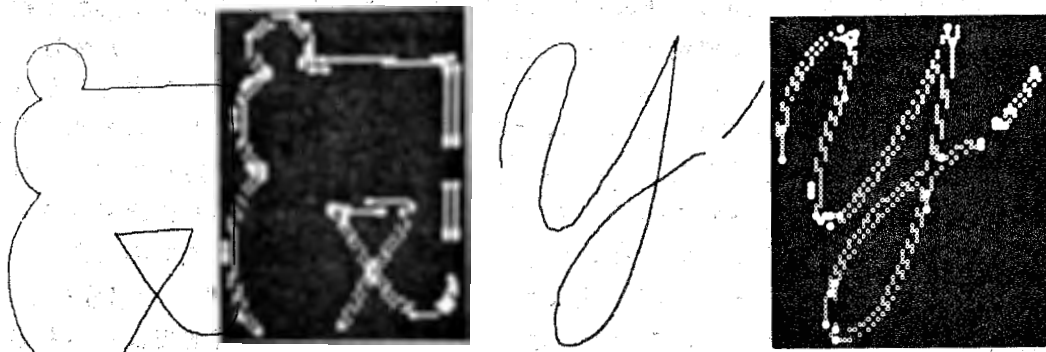


図6. 線図形を入力とした場合の特徴点

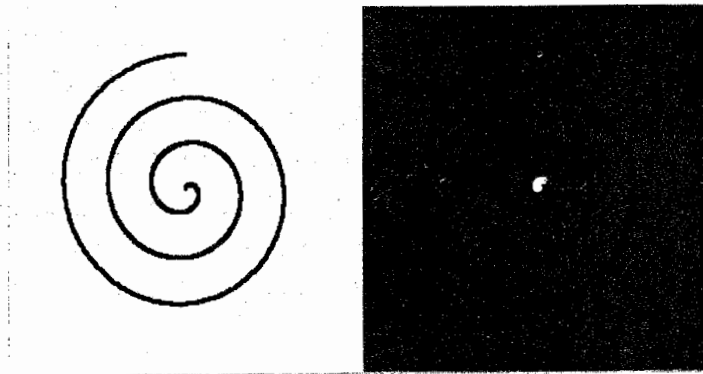


図7. 螺旋に対する特徴点

10.5 図形の構造化と特徴点検出

上記のようなアルゴリズムを用いれば、図形の特徴点を検出することができる。しかし、このアルゴリズムをそのまま用いれば、ノイズのような細かな凹凸をも検出してしまふ。我々は、まず、図形の大局的特徴となる点を検出する必要がある。そこで、図形の輪郭線を平滑化すると同時に、特徴点が検出できるようなアルゴリズムを考えた。これは、 $\nabla^2 G$ (Marr-Hildreth operator)⁽¹¹⁾ で画像を平滑化しゼロ交差をもとめた後に、単純型細胞様のフィルターを用いて上記のアルゴリズムと同様に特徴点を検出するものである。いま

$$G(r) = \frac{1}{2\pi\sigma^2} \exp(-r^2/2\sigma^2)$$

とすれば、

$$\nabla^2 G = (r^2/2\sigma^2 - 1) (\exp(-r^2/2\sigma^2)) / \pi\sigma^4$$

となる (図8)。

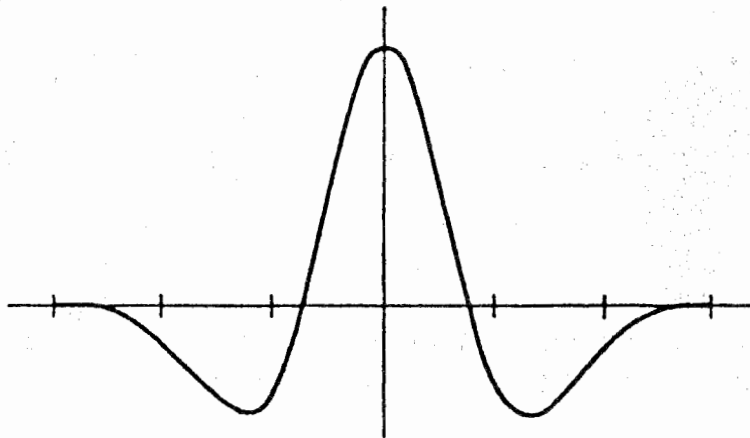
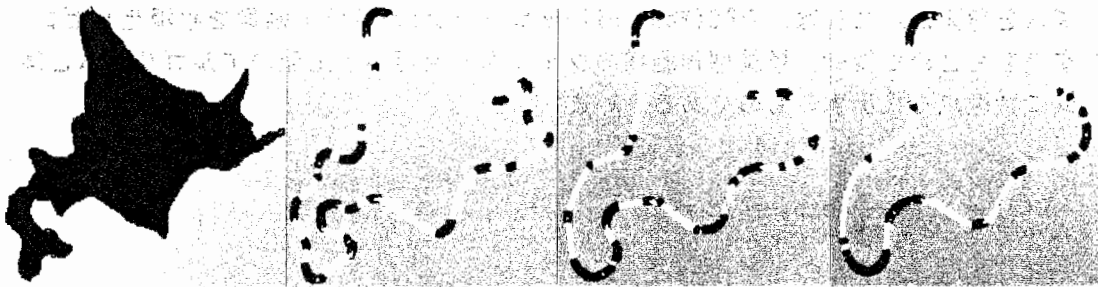


図8. Marr-Hildreth オペレータの1次元断面

ここで画像を $I(x,y)$ とすれば、 $\nabla^2 G * I = 0$ なる $I(x,y)$ がもつめるゼロ交差である。 σ の大きさを変化させることにより様々な平滑化ができ、解像度を変化させて特徴点を検出することが可能である。ここでは、まず画像のゼロ交差を求めた後、その出力に対して以下のような線検出器を作用させ前述の特徴点検出アルゴリズムと同様の方法で特徴点を検出した。具体的には σ は5, 10, 15, 20画素の4種類である。ゼロ交差を求めた後、 3×31 のマスクをかけ、その中に入るゼロ交差の数を計算する。この計算は各ゼロ交差点を中心に36方位で行ない、この最大値が25以下のゼロ交差点を特徴点としてマークした。

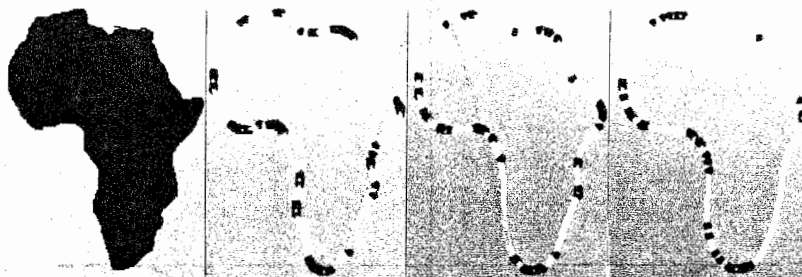
10.6 並列計算機NCUBEによるシミュレーション (特徴点検出)

256×256画素に対して1×1のサンプリング間隔で処理を行った。既に述べたようにノード数は512個なので、画像の1行分を2分割し、2個のノードに割りつけて処理を実行した。したがって各ノードは127画素を36方位で処理を行なった。計算時間は約2秒であった。北海道とアフリカ大陸に対する結果を図9および図10に示す。



左から、原画像、 $\sigma=10.0$ の結果、 $\sigma=15.0$ の結果、 $\sigma=20.0$ の結果。それぞれエッジの数は、766、678、642であり、特徴点の数は、182、129、110個であった。

図9. 北海道を入力図形とした場合の特徴点



左から原画像、 $\sigma=10.0$ の結果、 $\sigma=15.0$ の結果、 $\sigma=20.0$ の結果。それぞれエッジの数は、556、534、526であり、特徴点の数は、107、60、52個であった。

図10. アフリカ大陸を入力図形とした場合の特徴点

本アルゴリズムにおいても端点等の特徴点が抽出される(図1-1)。次節ではこれらの特徴点に基づき原画像の復元を試みる。

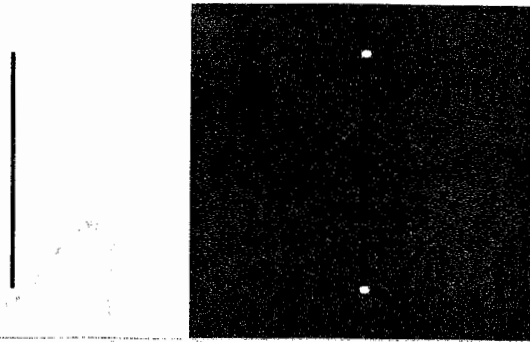


図1-1. 線図形を入力とした場合の特徴点

10.7 補間と画像復元

このような特徴がいったん検出されると、これをもとに図形を復元できる可能性がある。スプライン補間 C^1 級曲線では曲率がある程度不連続になると「不連続」が知覚される。しかし、 C^2 級曲線ではすべてなめらかに知覚されるといわれる。

一方、数学で「最もなめらかな曲線」とは、 $f(x)$ が k 回微分可能であるとき

$$\int_b^a [f(x)^{(k)}]^2 dx$$

を最小にする曲線のことを言い、これは、 $2k-1$ 次のスプライン関数になっている。そこで C^2 級で最もなめらかな曲線であり、minimum curvature property をもつ3次のスプライン補間をここでは考える。3次スプラインとは各節点(knot)間を異なる3次の多項式で補間するものである。

(1) 点線の補完

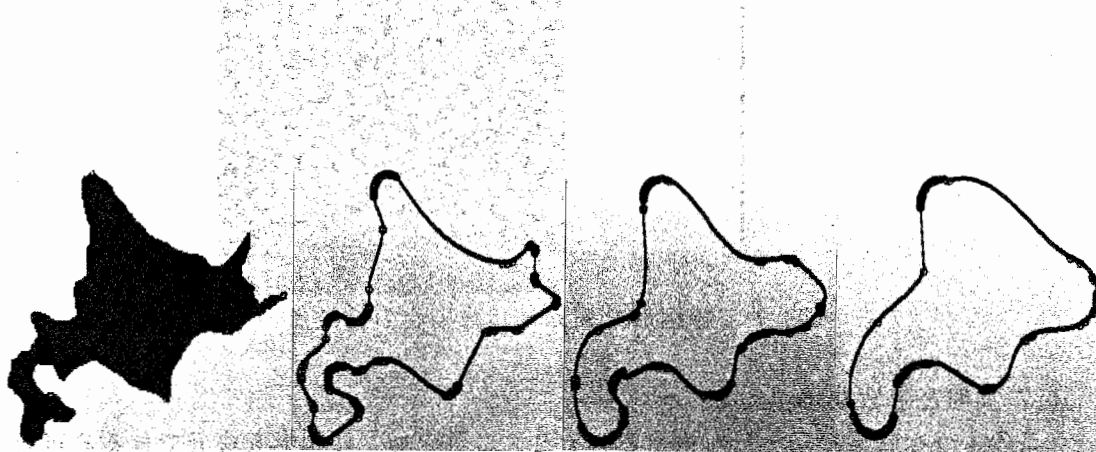
点が適当な条件下で配置されると曲線(直線ではなく)が知覚される。この問題を最初に論じたのは Koffka(1935)であろう。Koffka の例では、円周上に点が等間隔に配列されている。この場合、8点以下では8角形に見えてしまう。すなわち、直線補完がなされる。9点以上になると円が見えてくる。この場合3点がなす角度は 140° になる。SmitsとVos⁽¹²⁾は同様の実験を行い3点のなす角度が 140° 以上のときに曲線が知覚されるという結果を得ている。

(2) 補間アルゴリズム

ここでは、上記の知見を踏まえて特徴点から以下のようにして補間を行い原画像を復元する。隣合う2つの特徴点のなす角が 140° 以下のときは折れ線補間を、それ以外のときはスプライン補間を行う。

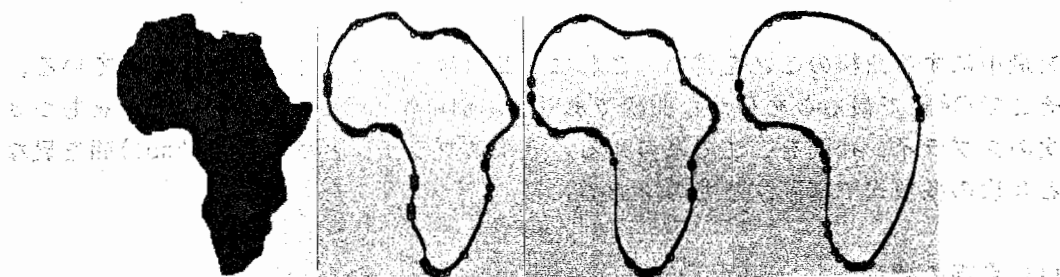
10.8 並列計算機NCUBEによるシミュレーション(特徴点からの画像復元)

上記のアルゴリズムに基づいて、図9および図10で得られた特徴点から原画像(図9、10の白線)の復元を試みた。結果を図12および13に示す。図から明らかかなようにかなり正確に復元された。



左から $\sigma=10.0$ の結果, $\sigma=15.0$ の結果, $\sigma=20.0$ の結果.それぞれ図10に対応する.

図12. 北海道の特徴点からの復元



左から原画, $\sigma=10.0$ の結果, $\sigma=15.0$ の結果, $\sigma=20.0$ の結果.それぞれ図11に対応.

図13. アフリカ大陸の特徴点からの復元

10.9 考察

2次元曲線の重要な特徴として、曲率の曲大点 (curvature extrema)、曲率の交差点 (cross point)、曲線の端点 (termination)、間隙 (gap) 等が挙げられる。心理学的研究から、我々はどこかに注意を向ける前に (すなわち前注意過程において) このような特徴をすばやく並列に処理していることが知られている。今回このような現象、すなわち異なる種類の特徴を持つ位置 (臨界点) を単一のアルゴリズムでシミュレートすることができた。本アルゴリズムから曲線の細部の情報は、高密度で存在する方向選択性の低い微小なチャンネルによって処理され、方向選択性の高いチャンネルは線条性の判定に用いられており、方向選択性の高いチャンネ

ルから方向選択性の低い微小なチャンネルへの抑制性の結合の存在が予想される。これらの3つの仮定を脳の神経系と対応させるなら、線検出器は4層あるいは6層のニューロンに、ゼロ交差は4と6層のニューロンに対応しているのかもしれない。最近6層から4層への抑制性のフィードバックがあることが報告されている(3)。

通常パターン認識の分野で使われている手法はあらかじめ特徴空間を与えておき、学習によって特徴空間の各軸の重みを変えていくという方式をとっている。一方細線化などの手法を用いて、パターンの骨格を出し骨格の交点や角、端点等の関係を抽出するという方法がある。しかしこの方法においては、このような特徴点を抽出することがむずかしいため、主として前者の方法がとられてきた。ここで示した並列アルゴリズムは、このような難点を克服するものであると考えられ、従来のパターン認識の前処理の部分に採用することによって、パターン認識能力を向上させることが期待できる。

10.10 文献

- (1) Marr, D.: Vision, Freeman(1987)
- (2) Zucker, S.W.: Early processes for orientation selection and grouping. Computer Vision Graphics Image Process(1986)
- (3) Dobbins, A., Zucker, S.W., and Cynader, M.S.: Endstopped neurons in the visual cortex as a substrate for calculating curvature, Nature, 329, pp. 438-441(1987)
- (4) Hubel, D.H., Livingstone, M.S.: Segregation of form, color, and stereopsis in primate area 18, The Journal of Neuroscience, 7, pp. 3378-3415 (1987)
- (5) Asada, H. and Brady, M.: The curvature primal sketch, IEEE Trans. P. A. M. I., 8, pp. 2-14(1986)
- (6) Freeman, H.: Shape description via the use of critical points, Pattern Recognition, 10, pp. 159-166(1981)
- (7) Attneave, F.: Some informational aspects of visual perception, Psychol. Rev., 61, pp. 183-193(1954)
- (8) Hoffman, D.D., and Richards, W.A.: Representing smooth plane curves for recognition: Implications for figure-ground reversal, in Proc. Nat. Conf. Artif. Intell., Pittsburgh, PA, pp. 5-8(1982)
- (9) Julesz, B.: Texton gradients: The texton theory revisited. Biological Cybernetics, 54, pp. 245-251(1986)
- (10) Canny, J.: A computational approach to edge detection. IEEE Trans., P. A. M. I., 8, pp. 679-698(1986)
- (11) Marr, D., Hildreth, E.: Theory of edge detection, Proc. R. Soc. Lond., B207, pp. 187-217(1980)
- (12) Smits, J. T. S. and Vos, P. G.: The perception of continuous curves in dot stimuli, Perception 16, pp. 121-131(1987)

1.1. エネルギー学習をする局所並列確率マルコフモデルを用いた輪郭線抽出(11), (12)

11.1 はじめに

輪郭線の抽出は、図形認識の基礎となる最も大切なものの一つである。画像をまず線図形に直し、それを用いて構造記述することが便利でかつ優れているために、輪郭線抽出がしばしば行われている(1), (2), (3)。一般にビデオカメラを用いて入力された「顔」等の自然画像の場合には、コントラストの変化が大きく、また反射や影等のノイズもあり、正確な輪郭線を抽出することは困難である。そこで、対象画像の形状に関する知識を用いることが必要になってくる。

コンピュータに形状の知識をもたせる際、対象画像の形状を完全な輪郭線として与えたのでは、対象画像の数だけ形状の知識が必要となるため効率が悪く、さらに知識を持たない対象の輪郭線は抽出できないことになる。そこで線画のごく近傍だけに着目して、端点や交点などの非常に基本的な近傍系の表現素を組合せることによって形状を表現し、表現素の出現確率として形状の知識を表現することが記憶効率、および知識表現の柔軟性の点で優れていると思われる。

Geman & Geman らによって、画像の形状の知識を確率場のパラメータとして表現し、このパラメータとノイズの乗った不完全な画像から原画像を復元する試みが行われている(4)。彼らは、ラインプロセスと呼ばれる画像の不連続性を表現する隠れた確率変数を用いて形状の知識を確率により表現し、シミュレーテッドアニーリング等の手法を用いて原画像の復元を行っている。また、D. Geman は、ラインプロセスが画像の濃淡値の不連続、すなわち輪郭線を表すことに着目して、おもちゃや家等の自然画像の輪郭線抽出を行っている(5)。ラインプロセスは、近傍系における連結の形式を、その出現確率を反映する局所的エネルギーとして持つことができるため、輪郭線の途切れを禁止するといった局所的な知識を持たせることができ、連結性に優れた輪郭線抽出を行わせることができる。さらに、Kawatoらはそれまでの手法では予め設定しなければならなかった局所的パラメータを学習により自動的に獲得する手法(6)を提案している。

本稿では、まず、確率緩和やシミュレーテッドアニーリングを用いて近傍系の表現素の出現確率として与えられた確率パラメータから輪郭線抽出を行う手法について述べる。次に、対象とする濃淡画像と輪郭線とを繰り返し学習させることで確率パラメータを自動的に調節する手法について述べる。さらに、これらの手法を用いて、人間の顔画像の輪郭線を抽出する方法および実行結果について述べる。最後に、従来のMRFモデルを拡張した階層的MRFモデルを提案する。

11.2 輪郭抽出のマルコフ確率場モデル

2次元のMRFを考える。まず S を画素およびラインプロセスからなる画像の格子点(サイト)の集合とする。また S の各要素 s に付随している確率変数を x_s で、さらに S 上の確率場を $X=(x_s, s \in S)$ で表す。 x_s の値域は、0 から $L-1$ までの離散 L 値であるとし、この値域を $\Lambda=(0, 1, \dots, L-1)$ で表す。それぞれの x_s が特定の値をとったものを状態配置 ω と表す。この ω がひとつの画像に対応する。 S

上の確率場 X がマルコフ的であるとき、全ての x_s に対する同時確率分布、すなわち画像 ω が生起する確率 $\Pi(\omega)$ は、次式で表される。

$$\Pi(\omega) = \frac{\exp(-U(\omega)/T)}{Z} \quad (1)$$

サイトを表す番号の系列 $\{n_t, t \geq 1\}$ が任意のサイトを無限に多く含むと仮定すると、任意の $X(0) = \eta$ について、確率変数 X の値の分布関数 $P(X)$ と同時確率分布関数 $\Pi(X)$ との間に、

$$\lim_{t \rightarrow \infty} P\{X(t) = \alpha \mid X(0) = \eta\} = \Pi(\alpha), \quad \forall \alpha \in \omega \quad (2)$$

が成り立つ。さらに、 S の近傍系が作るクリークを C 、クリークが持つ局所的エネルギーを V_C で表し、サイト s の確率変数が x でありかつ他のサイトが ω に一致する様な状態配置を ω^x で表すと、 $\Pi(\omega) = P(X = \omega)$ がギブス的であるならば、

$$P(X_s = x_s \mid X_r = x_r, r \neq s) = Z_s^{-1} \exp\{T^{-1} \cdot \sum_{C: s \in C} V_C(\omega)\} \quad (3)$$

$$Z_s = \sum_{x \in \Lambda} \exp\{T^{-1} \cdot \sum_{C: s \in C} V_C(\omega^x)\} \quad (4)$$

が成り立つ。 s に関するクリークの中の任意のサイトは s の近傍でなければならぬから、サイト s が x_s をとる確率は s の近傍の確率変数のみに着目した局所演算によって求めることができる。

したがって、各々のサイトにおける x_s の各々の値の出現確率 $P(\omega)$ を次式で定めて十分に状態変化を繰り返せば、局所並列演算によって同時確率分布 $\Pi(\omega)$ を実現することができる。

$$P(\omega) = \frac{\exp\{-U(\omega)/T\}}{Z} \quad (5)$$

ここで U は MRF のポテンシャルエネルギーであり、 S の近傍系が作るクリーク全体の集合 C の要素 C に含まれる ω のみに依存する局所的エネルギー V_C の和として表される。

$$U(\omega) = \sum_{C \in C} V_C(\omega) \quad (6)$$

また、 Z は分配関数と呼ばれる正規化のための定数であり、(7)式で表される。

$$Z = \sum_{\omega} \exp(-U(\omega)/T) \quad (7)$$

また、 T は温度と呼ばれるMRFの動作環境パラメータであり、 T の値によって各 ω の $P(\omega)$ の比をコントロールすることができる。この性質により 次の条件式を満足するスケジュールに従って T を減少させれば、状態の確率分布が $U(\omega)$ の最小値を与える状態については一様でそれ以外の状態については確率 0 とするような $\Pi(\omega)$ を実現することができる。

$$a) T(t) \rightarrow 0 \quad t \rightarrow \infty \quad (8a)$$

$$b) T(t) > \frac{N\delta}{\log(t)} \quad \text{for } t > t_0 \quad (8b)$$

ここで N はサイトの数、 δ は $U(\omega)$ の最大値と最小値との差である。

ポテンシャルエネルギーの定義式からわかるように、あるサイズのクリーク C が取り得る可能な全ての状態 C_i ($i=1,2,\dots,M$) について局所的エネルギー V_{C_i} の値を決めてやれば、それはマルコフ確率場モデルを、ひいては画像がこうあるべきだという事前知識を決定することになる。その上で、シミュレーテッドアニーリングを施しながら局所並列繰り返し演算を行えば $U(\omega)$ の最小値を厳密に求めることができる⁽⁴⁾。

画像の濃淡値を入力画像に固定したまま、ラインプロセスのみを変化させて輪郭線を抽出することを考える。いま、図1の様な近傍系を考えてみよう。

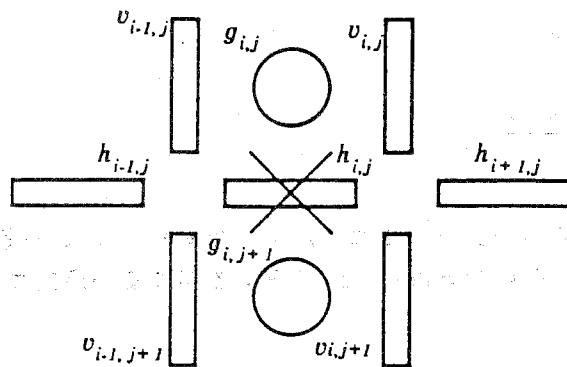


図1. 近傍系

|, -, ○はそれぞれ縦, 横のラインプロセスと画像の濃淡値を表す。×印をつけたサイトを中心に考える。サイト s_{ij} が取り得る状態は、そこにエッジが存在

する ($l_{ij}=1$), または存在しない ($l_{ij}=0$) のいずれかである。そこで、画素の濃度勾配のポテンシャルを $U_0(\omega)$, ラインプロセスのポテンシャルを $U_1(\omega)$, サイト全体のポテンシャルを $U(\omega)$ とするとそれぞれ次式のようなになる。

$$U(\omega) = U_0(\omega) + U_1(\omega) \quad (9)$$

$$U_0(\omega) = V_0 \cdot \sum_{i,j} \{ (g_{i,j+1} - g_{i,j})^2 \cdot (1 - h_{i,j}) + (g_{i+1,j} - g_{i,j})^2 \cdot (1 - v_{i,j}) \} \quad (10)$$

$$U_1(\omega) = \sum_{i,j} \sum_{C \in C} V_C(\omega) \quad (11)$$

このポテンシャルに従って十分に状態変化を繰り返すことによって、画像の境界として輪郭線を求めることができる⁽⁵⁾。

11.3 エネルギー学習

画像のマルコフ確率場モデルを特徴付ける局所的エネルギーを真の画像の状態を教師信号として同じ統計的性質を持つ画像データに対して推定・学習するアルゴリズム⁽⁶⁾が Kawato らによって提案されている。ここでは目標とする輪郭線画像を教師信号として濃淡画像から輪郭線画像を推定するのに必要となる局所的エネルギー V_{C_i} の値を学習するアルゴリズムについて説明する。

いま、対象としている輪郭線画像で、ある状態配置 ω が生起する確率を $P(\omega)$ とする。一方、局所的エネルギー V_{C_i} ($i=1, 2, \dots, M$) が決められたマルコフ確率場モデルが、同じ状態 ω をとる定常確率を $P'(\omega, V_C)$ とする。このとき、カルバックの divergence $G(V_C)$ は、次式のように定義される。

$$G(V_C) = \sum_{\omega} P(\omega) \log \{ P(\omega) / P'(\omega, V_C) \} \quad (12)$$

$G(V_C)$ は、 $P(\omega) = P'(\omega, V_C)$ のとき 0 で、それ以外では正である。マルコフ確率場モデルが輪郭線画像のなるべくよいモデルになるよう、 G の最急降下方向に局所的エネルギーを変化させるとすると、Boltzmann-machine の学習則の導出と同様にして次の学習則が得られる。

$$\begin{aligned} \Delta V_{C_i} &= -\epsilon \partial G / \partial V_{C_i} \\ &= -\eta \left(\sum_{C \in C} I_i(C) - \sum_{C \in C} I_i(C)' \right) \end{aligned} \quad (13)$$

ここで I_i はクリーク C の状態が C_i のときだけ 1 をとり、それ以外では 0 となる C_i の定義関数である。右辺第 1 項は、学習の目標となる輪郭線画像 (教師画像) において、クリーク C が C_i をとる個数を画像全体で数え、それをサイトの

数で平均したものである。一方右辺第2項は、ある局所的エネルギー V_{Ci} を固定した時のマルコフ確率場モデルで、クリーク C が C_i をとる個数の平均値である。これは、マルコフ確率場モデルで画像を生成して計算する。

11.4 顔画像輪郭抽出の実験結果

11.4.1 実験条件とパラメータ

縦128画素/横128画素、モノクロ256階調に表現された人間の顔写真の輪郭線を抽出した。原画像を図2(a)に、また V_{Ci} の学習を行う際に学習の目標とする輪郭線画像を同図(b)に示す。



(a)原画像 (b)教師画像

図2. 原画像と学習目標

ここでは11.2節で述べた1近傍のクリークに加えて平行二重線のコントロールを目的に "parallel" を加えて、図3に示す合計7つのタイプの近傍系を用いた。エネルギー学習をしない場合には、ラインプロセスのポテンシャルを求める際の係数 V_{Ci} は、図示の値を用いた。また、エネルギー学習をする場合には、この値を初期値として学習を行わせた。また濃度勾配のポテンシャルを求める際の係数 V_0 は、297.0とした。

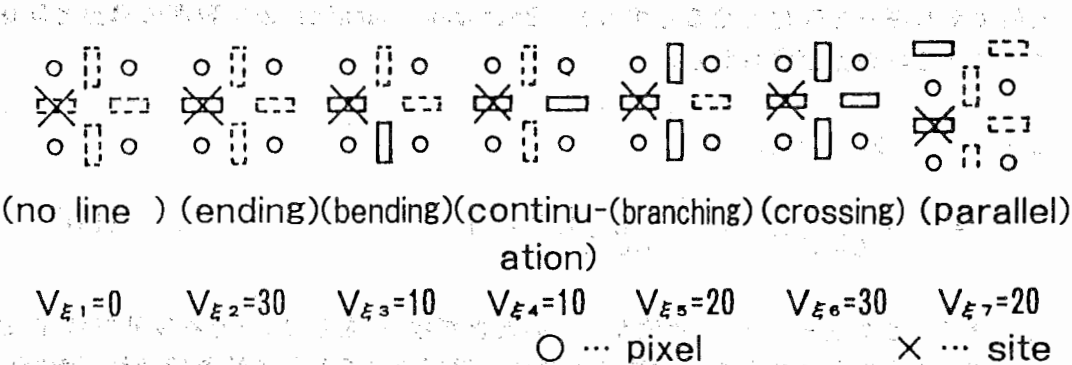


図3. 実験に用いた近傍系

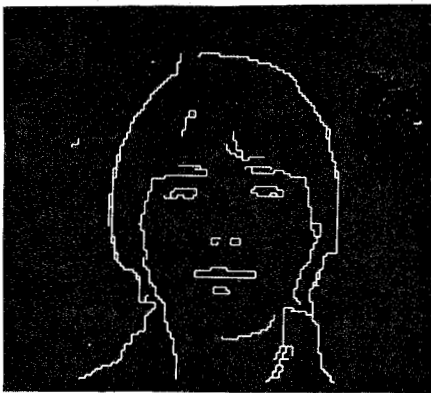
11.4.2 アニーリングスケジュール

金属結晶を作る際にアニーリングスケジュールが速すぎると結晶の中に傷が生ずるように、シミュレーテッドアニーリングにおいても、急激に T を下げると、マルコフ確率場の中に結晶の傷に似た雑音が生ずる。11.2 節で説明したアニーリングスケジュールを用いれば最適な同時確率分布が得られるが、多大な時間を必要とする。そこで次式のように定数 α を用いて α 倍の加速を行い、試行錯誤により傷ができない範囲の α と初期温度 T_0 を決定した。なお、アニーリングスケジュールを決める際には、エネルギー学習は行っていない。

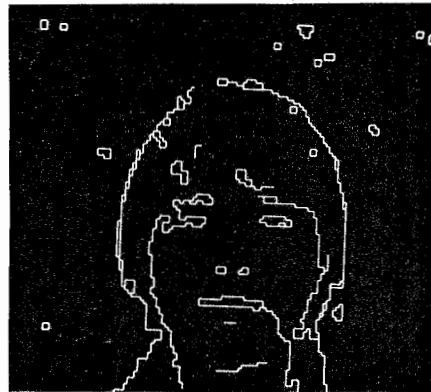
$$T(k) = \frac{T_0}{\ln(\alpha k + 2)} \quad (k : \text{sweep number}) \quad (14)$$

(1) 初期温度 T_0

まず $\alpha=1$ とし、様々な初期温度 T_0 について輪郭線の推定を行わせてみた。図 4 (a) は、 $T_0=50$ として約 4000 回 状態変化を行わせたものであり、(b) は、 $T_0=10$ として約 8000 回 状態変化を行わせたものである。(a) には見られない雑音が (b) に現れてきている。これは、 $T_0=10$ では初期温度が低すぎたためである。従ってここでは、初期温度 T_0 の値は 50 が適当であると考えられる。



(a) $T_0=50$



(b) $T_0=10$

図 4. 初期温度 T_0 を変えた場合の輪郭画像

(2) 加速係数 α

$T_0=50$ とし、様々な加速係数 α について輪郭線の推定を行わせた。図 5 (a) は $\alpha=10$, (b) は $\alpha=100$, (c) は $\alpha=1000$ として、それぞれ 1000 回づつ状態変化を行わせたものである。(a), (b) には見られない雑音が (c) に現れてきている。これは、 $\alpha=1000$ では加速が急すぎたためである。従ってここでは、加速係数 α の値は 100 が適当であると考えられる。

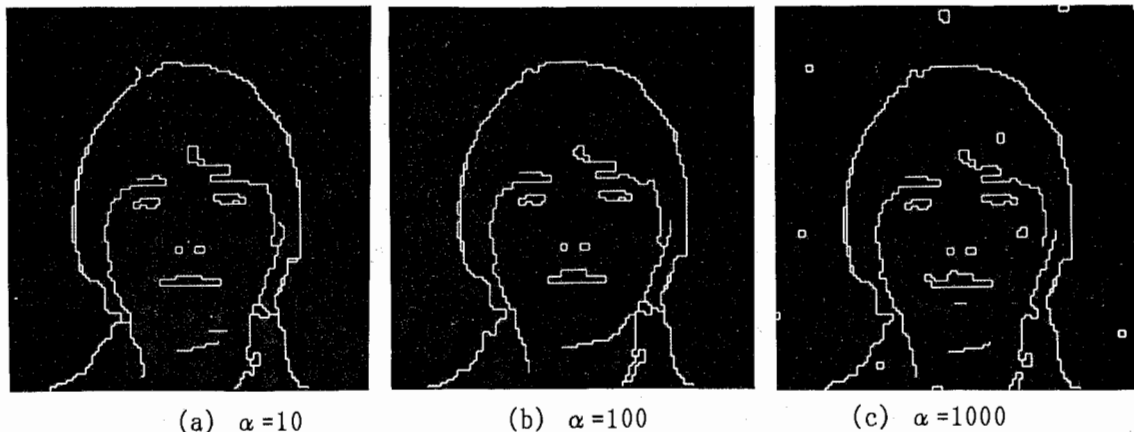


図5. 加速係数 α を変えた場合の輪郭画像

11.4.3 学習

第11.2節で説明した学習則に従えば、状態変化を十分に行った後に、マルコフ確率場モデルの中でクリーク C が C_i をとる個数の総和をサイトの数について平均したものと、教師画像中でクリーク C が C_i をとる個数の総和をサイトの数について平均したものの差に、微小な係数 η を乗じた値を ΔV_{C_i} として学習を行う。いま、 $\eta / (\text{サイト数})$ を改めて η とすれば、 C_i をサイト数で平均する必要はないので個数をそのまま用いることにする。

学習が進むにつれて、教師画像中の C_i の数とマルコフ確率場モデル中の C_i の数との差

$$\Delta N_i(c) = \sum_{C \in \mathcal{C}} I_i(C) - \sum_{C \in \mathcal{C}} I_i(C)' \quad (15)$$

すなわちクリーク数の誤差は減少する。しかし教師画像中に数が少ない C_i においては、教師画像中の C_i の数に対するクリーク数の誤差の比率、すなわち

$$\frac{\Delta N_i(C)}{\sum_{C \in \mathcal{C}} I_i(C)} \quad (16)$$

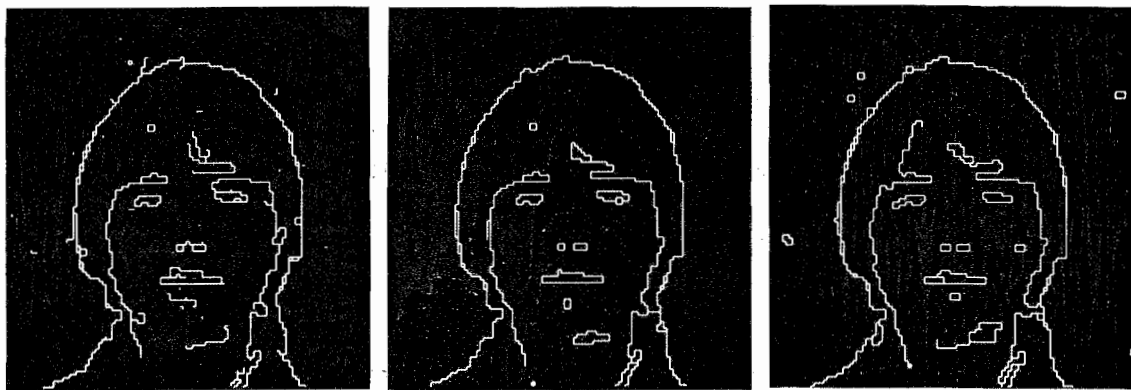
は、なかなか減少せず、エネルギー学習に非常に時間がかかる。そこで、教師画像とマルコフ確率場モデルとの C_i の数の差を次式のように教師画像中の C_i の個数で規格化したものを ΔV_{C_i} として学習を行った。

$$\Delta V_{C_i} = -\eta \left(\sum_{C \in \mathcal{C}} I_i(C) - \sum_{C \in \mathcal{C}} I_i(C)' \right) / \sum_{C \in \mathcal{C}} I_i(C) \quad (17)$$

なお、教師画像中の C_i の個数が 0、すなわち $\sum_{C \in \mathcal{C}} I_i(C) = 0$ の場合には ΔV_{C_i}

の値が定まらないので、規格化を行わずに ΔV_{c_i} とした。また、“No lines” は最小のエネルギーのままであることは明白なので、 V_{c_i} は学習を行わせなかった。また以下では、 η の値を試行錯誤により選定し、0.1 とした。

図6 (a) は学習前、(b) は1回学習後、(c) は2回学習後のエネルギーパラメータにより各々 10000回 の状態変化を行った結果得られた輪郭線画像である。(a) では端点部分 “ending” が非常に多くみられるが、(b) ではこれが著しく減少していることがわかる。さらに、(c) では前髪と顎の輪郭線が延びて連結しつつあることがわかる。しかし、“ending” や “parallel” 等を禁止する必要から、学習によってこれらにかかわる V_{c_i} が著しく増大したために $U(\omega)/T$ が増大し、結果的に T_0 が下がったことと等価となり、生成された輪郭線画像中に雑音が現れてきている。



(a) 学習前 (b) 1回学習 (c) 2回学習

図6. 学習による輪郭画像の変化

この問題点を解決するには、学習による V_{c_i} の増大を見込んで初期温度 T_0 の値を十分に高く設定することが必要となる。とくに、教師画像中に存在しない C_i (個数が 0 の C_i)がある場合は、 V_{c_i} が極めて大きくなるため、 T_0 を非常に高く設定することが必要になる。ここで用いた教師画像の場合には、その中に “ending” が存在していないため、学習により V_{c_2} が極めて大きくなることが予想される。したがって T_0 を非常に高くしてゆっくりとアニーリングすることが必要であり、多大な時間が必要となる。

11.4.4 Constrained Optimization

前節の問題点を克服する手段として、D.Geman 等が提案している Constrained Optimization という手法⁽⁷⁾がある。これはアニーリングによって温度 T を下げて行く際に、アニーリングスケジュールと一定の関係を保ちながら特定の V_{c_i} の値を強制的に増加させて行く手法である。すなわち 禁止したい C_i に対応する V_{c_i} のみに λ なる係数を乗じ、この λ を次の条件式を満足させるように増加させるというものである。

$$T^{-1} \cdot \lambda \leq \text{const} \cdot \ln(k) \quad (k : \text{sweep number}) \quad (18)$$

したがって、アニーリングスケジュールを

$$T(k) = \frac{T_0}{\ln\{\alpha(k-1)+e\}} \quad (19)$$

とすれば、 λ は

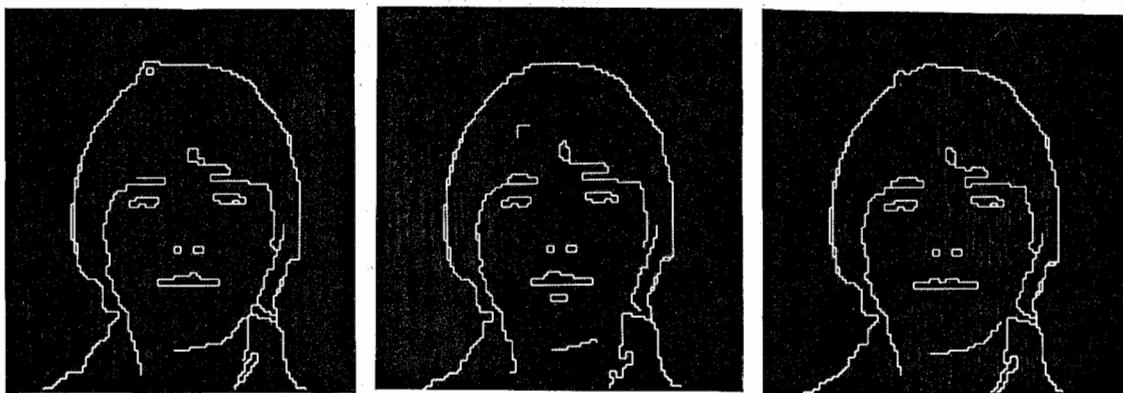
$$\lambda(k) \leq \frac{\text{const} \cdot T_0 \cdot \ln(k)}{\ln\{\alpha(k-1)+e\}} \quad (20)$$

となり、この条件式を満足する範囲で λ を増加させれば、 C_i を禁止することができる。

図7に、Constrained Optimizationを導入して抽出した輪郭線を示す。const=1/30, $T_0=50$, $\alpha=100$ とし、 λ を次式のように定めて輪郭線の抽出を行った。

$$\lambda(k) = \frac{\text{const} \cdot T_0 \cdot \ln(k)}{\ln\{\alpha(k-1)+e\}} \quad (21)$$

(a)は学習前、(b)は1回学習後、(c)は2回学習後のエネルギーパラメータにより各々1000回の状態変化を行った結果得られた輪郭線画像である。



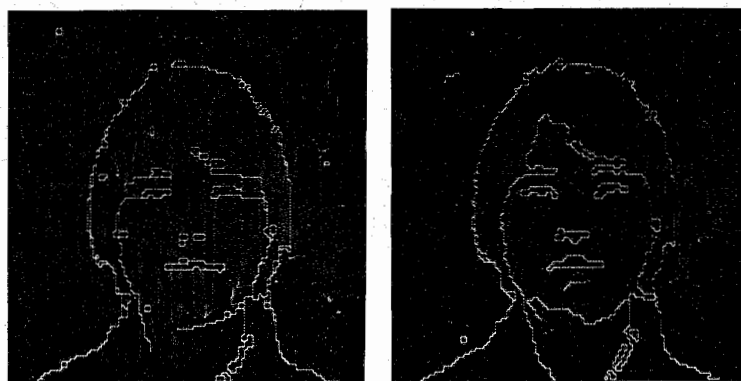
(a) 学習前 (b) 1回学習 (c) 2回学習

図7. Constrained Optimization を用いた輪郭画像

11.4.5 並列計算機NCUBEによる高速化

NCUBEは、1024個までのプロセッサをハイパーキューブ状に接続できる並列計算機である。シミュレーテッドアニーリングを施しながら繰り返し計算を行い、さらに学習を行うには多大な時間を必要とするが、MRFモデルは局所並列演算が可能であり、NCUBEを用いることにより劇的に計算時間を短縮することができる。

図8に11.4.4節と同じアルゴリズム(const=2/30)によって抽出した輪郭線を示す。画素数が128×128であることから、128個のプロセッサを用いることとし、各々に横1ラインの処理を割り当てた。各ラインへのノードの割付は、1次元鎖⁽⁸⁾の方法を用いた。VAX8650においては1000回の状態変化を行うのに約7600秒を要するが、NCUBEでは約710秒であった。したがって、NCUBEを用いることにより実行時間を約1/10以下に短縮することができた。



(a) 学習前

(b) 30回学習

図8. NCUBEを用いて抽出した輪郭画像

11.4.6 従来手法との比較

本手法で得られた輪郭線画像がどの程度の性能であるかを知る目的で、いくつかの輪郭線抽出法との比較を行ってみた。図9に各々の方法で抽出した輪郭線画像を示す。(b)のゼロ交差および(h)のOhtsuの方法以外は、閾値が必要なので、試行錯誤により輪郭線の連結性が保たれる範囲でできるだけ雑音が少ない点に設定した。なお、(b)以外は、通商産業省電子技術総合研究所によって開発されたグラフィックライブラリであるSPIDERを用いた。

(a)は3×3のLaplacianオペレータを用いて求めた輪郭線であるが、非常に連結性が悪い。(b)は6×6の ∇^2G フィルタをかけた後にゼロ交差を求めたもの⁽⁹⁾であるが、雑音が多い。(c)はHueckelオペレータを用いて求めた輪郭線であるが、連結性が良くない。(d)はKirschオペレータ、(e)はRobertsオペレータ、(f)はRobinsonオペレータ、(g)はSobelオペレータを用いて求

めた輪郭線であるが、閾値を調節しても雑音が消える前に輪郭線が切れてきている。(h)はOhtsuの方法により濃淡画像を2値化した後に、境界を追跡して輪郭線を求めたものであるが、雑音が少なく、かつ輪郭線の連結性も良好であり、比較的良好な輪郭線画像が得られている。(i)は11.4.4項において得られた輪郭線画像である。



図9. 従来手法により求めた輪郭画像

11.5 考察

11.4.5項で行った比較からもわかるように、エネルギー学習をするMRFモデルによって良好な輪郭線が得られた。従来手法が単に濃度変化のみで輪郭線を求め

ているのに対して、MRFモデルはさらに輪郭線同士の相互作用までも操作することができ、平行二重線 (Parallel) や端点 (ending) 等を必要に応じて禁止したりできるという優れた特徴を持っている。この性質を用いて端点や平行二重線を制限した輪郭線検出を行うことができた。

しかしMRFモデルには、直接的には対象画像の全体的な構造は無く、局所的状態変化の繰り返しの結果としての間接的な全体しか存在しない。したがって、得られる輪郭線画像は、ある時点で各クリークにとって最も都合の良い状態の集合でしかなく、ひとつの構造としての"輪郭線"には必ずしもなっていない。いわばこのような局所的な拘束条件だけを持ち全体的な拘束条件(構造)を持たないところにひとつの限界があると思われる。そこで我々は、画素数、すなわち"画像を表現する粗さ"が異なる複数のレベルを設け、その各々に対してMRFモデルを適用し、さらに各レベル間に相互作用を有する"階層的MRFモデル"を提案する。

図10に階層的MRFモデルの概念図を示す。(a)は階層間の対応を表す概念図である。粗いレベルにおいては1つの画素が、細かいレベルにおいて同じ位置にある複数の画素に対応する。図の例では、レベル $n+1$ の1画素がレベル n の4画素に対応している。一方、(b)は階層間のクリークの対応の例である。先に粗いレベル(level $n+1$)で輪郭線を求めておき、それより1段細かいレベル(level n)で輪郭線を求める際に、level $n+1$ でのクリークの要素と対応するクリーク要素のエネルギーが低くなるように条件づけて計算を行う。すなわち、0から $N-1$ までの N 個のレベルからなる階層的MRFモデルにおいて、level n におけるクリークを $C_n = \{c_n, n=1, 2, \dots, N\}$ 、状態配置 ω に対してクリークの要素 c_n ごとに与えられる局所的エネルギーを $V_n^*(\omega, c_n)$ 、1または微小な値 ϵ をとる係数を A とすると、実際にエネルギー計算に用いる局所的エネルギー $V_n(\omega, c_n)$ は次式のように表すことができる。

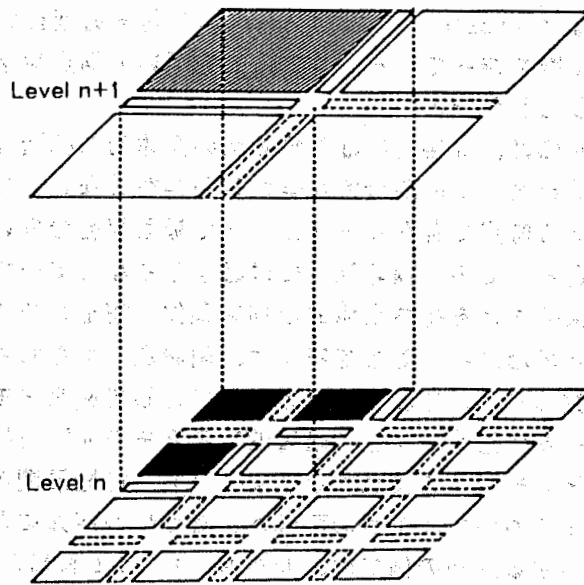
$$V_n(\omega, c_n) = V_n^*(\omega, c_n) \cdot A \quad (22)$$

$$A = \begin{cases} \epsilon \ll 1 & \text{if } c_n \in C_n \\ 1 & \text{otherwise} \end{cases}$$

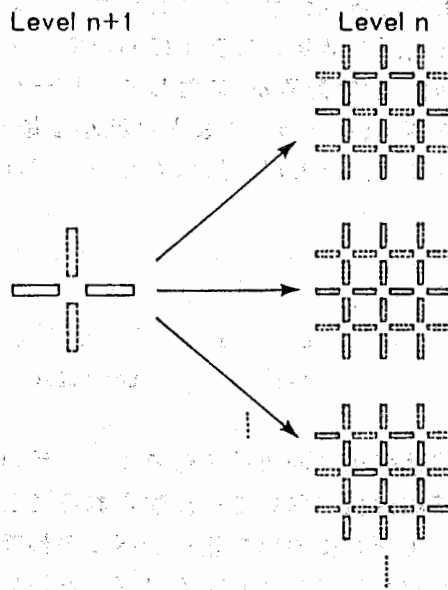
この手法を用いれば、状態変化の局所性を保ちつつ、粗いレベルで求めた全体的構造が細かいレベルの輪郭線画像においても反映されることになる。

さらに、階層的MRFモデルにおいて粗いレベルの輪郭線を求める際に T. Inui 等が提案している"特徴点検出アルゴリズム"⁽¹⁰⁾を用いることによって、より正確な輪郭線を求めることができると考えられる。このアルゴリズムは、複数のサイズのエッジオペレータを組み合わせて輪郭線で曲率の大きな部分、分岐点、端点、交点等を同時に抽出するものである。このアルゴリズムによって特徴点を検出し、特徴点の間をMRFモデルを用いて補間して行くのである。具体的には、特徴点が検出されたサイトは特徴点の形状に対応するラインプロセスの配置のポテンシャルを下げておく。その上で"continuation"のエネルギーを低くし、逆に"ending"のエネルギーを高くしてやれば、特徴点間をできるだけ直線的に補間することがで

きるものと考えられる。



(a) 階層間の対応



(b) クラックの対応の例

図10. 階層的MRFモデルの概念図

11.6 おわりに

本章では、エネルギー学習をするMRFモデルを用いて自然画像の輪郭線を抽出する試みについて報告した。実験の結果、シミュレーテッドアニーリング、エネルギー学習、Constrained optimization等の手法を用いて良好な輪郭線画像が得ら

れた。さらに画像の全体的構造を反映する手法として "階層的MRFモデル" を提案した。階層的MRFモデルの実証が今後の課題である。

11.7 文献

- (1) L.S.Davis: "A survey of edge detection techniques", Computer Vision, Graphics, and Image processing, Vol.4, pp.248-270(1975)
- (2) R.M.Haralick: "Edge and region analysis for digital image data", Computer Vision, Graphics, and Image processing, Vol.12,pp.60-73(1980)
- (3) P.K.Sahoo and et. al.: "A survey of thresholding Techniques", Computer Vision, Graphics, and Image processing, Vol.41,pp.233-160,(1988)
- (4) S.Geman and D.Geman: "Stochastic relaxation, Gibbs distributions and the basian restoration of images", IEEE Trans., Vol. PAMI-6,pp.721-741 (1984)
- (5) D.Geman: "Stochastic model for boundary detection", Image and vision computing, Vol.5, No.2, pp.61-65(1987)
- (6) M.Kawato and et. al.: "Information processing of image and neural network models", J. of Japanese society for artificial intelligence, Vol.4, No.2(1989)
- (7) D.Geman and et.al.:"Boundary detection by constrained optimization", preprint(1988)
- (8) T.Ikeda and et. al.: "Energy learning in neural network model which reconstructs image from noizy data", ITEJ Technical Report, Vol.12; No.14, pp.31-36(1988)
- (9) D.Marr: "Vision", Freeman(1982)
- (10) T.Inui and S.Miyake: "Structural shape description and a model of human visual memory",IEC Technical Report,Vol.MBE88-107,pp.91-98(1989)
- (11) 本郷,川人,乾,三宅:輪郭線抽出をする局所並列確率アルゴリズムのエネルギー学習,電子情報通信学会春季全国大会(1989.3.29 発表予定)
- (12) 本郷,川人,乾,三宅:エネルギー学習をする局所並列確率アルゴリズムを用いた輪郭線抽出,電子情報通信学会MEとバイオサイバネティクス研究会 (1989.3.15 発表予定)

12. おわりに

A T Rにおける視覚・認知機構研究における並列処理計算機NCUBEの利用について具体的事例を中心にまとめた。並列処理計算機、ことに疎結合型については従来の計算機利用とは異なる発想が要求される。例えば、導入の際に種々の応用例やユーティリティが組み込まれているわけではなく、データや問題の並列性を有効に利用しようとするユーザー群によって徐々にそれら応用例が蓄積されて初めて汎用性を身につけていく。この意味でユーザー主導によってのみ存在価値が高まる新しい性格の商品であると言えよう。本稿によって並列的計算原理の有効性とそれを実際に動かす道具に関する認識が高まることを期待する。

視覚・認知機構研究においては従来の計算機でモデルをシミュレーションすることも十分可能であり、手中の道具立てを念頭においた研究に限界があることも事実である。しかしながら、画像や音声パターンといった実世界の超並列データを高速で扱える道具の供給は、単なる高速化という「研究の量的変化」だけでなく会話的、実時間的シミュレーションによる「研究の質的变化」にも大きく寄与する。基礎研究で培われた並列計算原理の実証が可能となり、マシンアーキテクチャにとらわれないさらに自由な発想に基づく原理やモデルの構築が促進されるからである。

NCUBEの利用は上述の問題意識を持たせてくれた最初の1ステップである。今後、生理学や心理学の実験、計算理論、計算機工学、デバイスなど様々な分野が協調し合う学際研究が有効に機能し、新たな成果が生み出されることを期待する。

国内1号機のNCUBEを利用した研究の最初の1歩を踏み出すにあたり大変多くの方々の方々の助力と支援がありました。ここに以下の方々に深甚なる謝意を表します。

NCUBE導入を支援し研究の機会を与えていただいたA T R視聴覚機構研究所葉原耕平会長、淀川英司社長、梅田三千雄前視覚研究室長（現N T T）に感謝いたします。NCUBE選定にあたり助言をいただいた早稲田大学村岡洋一教授、事前評価実験を助けていただいたNHK放送技術研究所伊藤崇之研究員に感謝します。第4章の研究では、名古屋大学鳥脇純一郎教授、安田孝美助手に御助力をいただきました。第5章の研究は、N T Tヒューマンインターフェース研究所高橋裕子研究員を中心に実施されたものであり同所塩昭夫主任研究員にも多大の助言を賜りました。第6章の研究には滋賀医科大学可児一孝教授の御助力がありました。第7章について研究の機会を与えていただき御助言をいただいた東京大学鈴木良次教授に深謝します。第8章はNHKとの共同研究によるものであり、尽力いただいたNHK放送技術研究所日下秀夫前部長、福島邦彦主任研究員に感謝します。第9章、10章、11章について有益な議論をいただいた認知機構研究室の研究員諸氏に厚く感謝致します。NCUBE納入時に環境整備と立ち上げを助けていただいた住商電子システム山上裕司氏に御礼申し上げます。第3章の一部について御助力いただいた視覚研究室城和貴研究員に感謝します。国外では、いつも助言をいただくカリフォルニア工科大学のProf. G. C. FoxならびにProf. J. Hopfieldに感謝致します。最後に、いつも研究のサポートをしていただく岸本洋一課長をはじめとする当所企画課の皆様感謝いたします。