

物体認識の神経機構と双方向処理モデル

安藤広志

ATR人間情報通信研究所

本論文では、3次元世界における物体認識の問題を認識手法と神経機構の両面から検討し、「物体像表現の学習」及び「双方向緩和計算」を特徴とする物体認識モデルを提示する。また、従来の認識モデルと比較した本モデルの特徴について述べるとともに、心理学・神経科学の最近の知見との関連について議論する。さらに、3次元物体の実濃淡画像を用いて行なった計算機実験に関して、本モデルによる機能の実現例を示す。

Neural Architectures and Bidirectional Model for Object Recognition

Hiroshi Ando

ATR Human Information Processing Research Laboratories

This paper discusses computational and biological issues on object recognition in the 3D world, and presents an object recognition model that is based on view learning and bidirectional relaxation. The paper also compares the bidirectional approach with the conventional unidirectional approaches, and describes its relation with some recent findings in psychology and physiology. The performance of the model is further evaluated through computer experiments that use gray-level images of several 3D objects.

1 はじめに

我々人間は、3次元世界における物体の認識をいとも簡単に行なっているように思える。しかし、それに類した機能を計算機で実現しようとするとき必ずしも容易でないことがわかる。それは、画像上の多様な変化が、実は様々な要因の複合により生じているからであろう。まず第一の要因は、対象物体に応じてその形状やテクスチャが多様に変化することである。特に、顔や身体など非剛体物体が認識対象になる場合は、幾何学的な拘束を利用するのが難しくなる。第二に、視点変化または物体自体の姿勢変化により画像は著しく変化する。これは、奥行きの変化だけでなく物体の自己遮蔽により裏の形状や模様の情報が得られないことに起因する。第三に、複数の物体が存在する場合、全体的・部分的な遮蔽（重なり）が生じ得る。第四に、物体と視点との位置関係により、画像上で物体の位置や大きさに変化が生ずる。第五に、照明の変化により、影や陰影に変化が生じる。これらの基本要因に加えて、外部・内部のノイズにより情報に劣化が生じる。このような形状・テクスチャ・視点・遮蔽・位置・大きさ・照明などの変化に起因する変動が畳み込まれた画像から、どのようにすれば物体を認識できるのだろうか。このような物体認識の計算原理と計算機構を探ることは、神経科学と計算機科学の共通の課題であると言える。本稿では、「物体像表現の学習」及び「双方向緩和計算」を特徴とする物体認識モデルについて述べ、その機能的な意味や神経科学的な妥当性について議論し、計算機による機能の実現例を提示する（Ando, 1996; 藤田と安藤, 1997）。まず、以下に学習と双方向計算の基本概念について説明を行なう。

学習 (learning) は、一般的には広範囲の意味を含む概念であるが、数理的には多変量の統計処理と見なすことができる (Duda & Hart, 1973)。特に、線形だけでなく非線形まで含めた様々な多変量統計手法が展開されてきている (Hertz et al. 1991; Bishop, 1995)。例えば、教師あり学習 (supervised learning) の基本型は、関数近似 (function approximation) であり、入出力の例 (examples) から (非線形) 関数のパラメータを推定する方法である。

一旦パラメータが決まれば、学習時に用いていない入力値に対しても、関数の補間能力に応じて出力値が推定できる（汎化性）。特に、MLP（Multi-layer perceptron）やRBF（Radial basis function）などの非線形関数が一般に用いられる。これらの関数はネットワークで表現可能であり、十分なユニット数があれば任意の非線形関数を近似できることが証明されている（Funahashi, 1989）。一方、教師なし学習（unsupervised learning）は、出力を明示的に与えず、関数型や学習則の拘束だけでデータに潜む統計的構造を抽出する手法である。例えば、次元圧縮やクラスタリングがこれに相当する。学習は、従来からパターン認識の主要課題である（Duda & Hart, 1973）が、最近では特に初期視覚や物体認識の処理に積極的に用いられてきている（Nayar & Poggio, 1996）。これは、環境の物理的拘束に基づく手法に加えて、統計的な手法を用いて、より柔軟な課題達成を目指しているからである。特に、神経系においては、学習や適応が感覚・認知・運動全般において本質的な役割を果たしていると考えられる。

一方、双方向処理（bidirectional processing）とは単一方向処理に対する用語であり、処理を段階的に進めるのではなく、複数の処理系の間で情報を循環させる処理様式を指す。物体認識に限らず視覚の多くの課題は、一般に複数の要因に規定されているため、複数の処理系の持つ拘束を統合する双方向的な処理様式、すなわち統合による分析（analysis-by-synthesis）手法が適していると考えられる。特に、処理が階層構造を持つ場合は、上位（高次）の処理系の結果を下位（低次）の処理系にフィードバックして再処理を行なう。このとき上位の処理系からの情報は、下位の処理系に対する仮説や予測として機能する。例えば、物体認識の中間処理として画像分析や画像解釈を行なう場合、入力情報に加えて、記憶情報から生成された物体イメージを利用することにより、処理を促進できるものと考えられる。従ってこれは、「生成を用いた認識」手法とみなすことができる。

本稿では、まず2節において、従来の単一方向的な認識モデルを概観し、双方向処理の特徴を述べる。3節では、脳の認識機構に関する心理学的・生理学的知見を整理し、本アプローチとの関連を検討する。4節では、双方向処理を利用した物体認識をモデル化する試みについて述べる。このモデルは、混合型自己連想ネットワークの学習と双方向緩和計算による照合に基づいている。5節では、このモデルを用いた計算機実験の結果の一部を示す。6節では、今後に残された課題を考察した後、本稿を締めくくる。

2 物体認識モデルと双方向的アプローチ

本節では従来の物体認識モデルを概観し、双方向処理に基づくアプローチの特徴について述べる。まず、物体認識のボトムアップ的な手法としては、次のようなものがある。第一に、3次元表現素（3D primitives）を組み合わせて、対象物体の構造を物体中心座標系で記述する方法（Structural description approach）がある。例えば、一般化円錐（Agin & Binford, 1976; Marr & Nishihara, 1978）を組み合わせて物体を記述する方法が提案されている。また、3次元表現素を用いた心理学的な物体認識モデルとしてはGeon modelがある（Biederman, 1985）。これらの手法で物体構造を忠実に記述できれば有用であるが、3次元表現素を構築するための計算コストが高い点と、任意の形状が必ずしも3次元表現素に成分分割できない点に問題があると考えられる。第二は、幾何学的不変量（Geometrical invariance）の検出による認識手法である（Mundy & Zisserman, 1992 など）。この手法は数理的な厳密性に基づく点に特長があるが、実世界の多様な物体画像に適用して不変量を抽出するのは必ずしも容易ではないと思われる。第三の手法は、様々な視点から得た物体投影像を内部表現として用いる認識方法（View-based approach）である。UllmanとBasri (1991) は、少数の物体投影像（view）の線形結合で任意の投影像を表現できることを示し、この手法を物体認識に適用している。また、PoggioとEdelman (1990) は、複数の投影像を学習によりプロトタイプとして獲得し、それらの補間（汎化性）により視点不変性を実現するRBFネットワークモデルを提案している。さらに、主成分分析（KL展開）を用いて画像次元を圧縮し、低次元空間で3次元物体の識別を行なう手法が提案されている（Murase & Nayar, 1995）。これらView-basedの手法は、対象の形状を限定しないため適用の自由度が高く、処理が比較的単純であることが利点である。ただし一般に、物体や投影像の増加に必要な記憶量が増大する傾向があるため、記憶コストを削減する工夫が必要となる。

もう一つのアプローチは、トップダウン処理を重視した認識方式である（Lowe, 1987; Grimson, 1990等）。これは、仮説-検証型の手法であり、予め記憶しておいた物体のモデルや知識を初期の視覚計算に対して能動的に用いる方法（Model-based approach）である。この方法は、入力情報だけから画像を解釈するのが困難な状況、例えば、複数の物体間で遮蔽が生じているような複雑な情景の解析には、特に効力を発揮する。例えばLowe (1987) は、部分的に遮蔽された物体を認識するために、3次元的な物体モデルを用いて、入力物体像の視点方向を推定し、その視点から見た画像を物体モデルより生成して入力画像と照合を行なっている。ただし、この手法は視点

推定に幾何学的不変特徴（線分の結合性や平行性など）を用いているため、対象物体は直線で構成された多面体に限定されている。一方、WiskottとMalsburg (1993) はview-basedの手法を採っており、Gabor関数で抽出された特徴画像を記憶しておき、複数の物体が重なり合って存在する情景（視点方向は固定）に対して物体領域の検出と照合を行なっている。ただし、視点変化に対処するには、様々な視点に対応した多量の物体像と逐次的照合に必要な多量の計算が必要となるだろう。従って、特徴画像を単に記憶して入力画像に当てはめるだけでなく、どのような物体表現を獲得し、それらを入力に応じてどのように選択していくかをモデル化する必要があらう。

以上述べた認識手法に対して、双方向的な処理様式が考えられる。単一方向処理では複数の処理を逐次的に行なうのに対して、双方向処理では異なる処理系の間で情報を循環させる。この方式では、複数の処理系の間で仮説生成と仮説検証を繰り返すことにより、異なるレベルの拘束を融合し、全体的な精度を高めることを目標とする。特に、物体認識においては、入力情報から画像分析を行なう上昇経路と記憶情報から物体像生成を行なう下降経路を反復することにより両者の信頼性を高める。さらに、より柔軟なシステムを構成するためには、設計者が物体表現を予め完全に設定したり、逆に全ての画像を記憶するのは避け、課題達成に必要な物体表現を学習により獲得する方式が望ましい。4節では、このような様式を目指した一つの物体認識モデルについて述べる。

一方、これに類した双方向的な処理様式は他にも提案されてきている。まず、Mumford (1992) はフィードバックループを含む認識モデルを提案し、脳の認識機構としての妥当性を検討している。またUllman (1996) も、視覚皮質の神経機構と関連づけながら、双方向的な認識モデル (Counter-stream model) を主張している。Ullmanのモデルでは、時間のかかる繰り返し計算を避けるため、上向き・下向きの両方向から多重の表現を生成して並列的に照合を行なう考え方を提案している。ただし、何れのモデルも概念的な枠組みに留まっており、その具体的な実現様式は必ずしも明確ではない。また、RaoとBallard (1997) は、Kalman filterを用いて上位から下位の表現を能動的に予測する認識モデルを提案している。ただし、このモデルでは異なる視点からの物体表現の獲得は行っていない。さらに、選択的注意の神経回路モデルも提案されている (Olshausen et al., 1993など)。これらのモデルでは、上位から物体表現の生成を行なうというよりは、下位の表現を活性化 (抑制) したり、上方向の情報の流れの制御を行なっている。一方、物体認識以外の課題でも双方向の処理様式が提案されている。Kawatoらは、順逆緩和モデルを脳の計算モデルとして考案し、運動制御における軌道生成問題や視覚の再構成問題 (Kawato et al., 1993) に適用し、その有効性を示している。以上述べたように、双方向計算にも様々な形態が可能であり、それらの関連もまだ明確ではないが、今後の展開が期待される考え方と言えよう。

3. 神経系における物体表現と認識過程

本節では、ヒトやサル神経系における物体認識機構に関して、物体表現と認識過程という二つの観点から、これまで得られている知見を整理し、双方向モデルとの関連を議論する。まず、物体表現に関しては、ヒトの認識特性を心理物理実験で検討することにより、3次元物体の内部表現が視点不変か視点依存かという議論がなされてきた。まだ完全な決着はみていないが、基本カテゴリーの物体認識 (机と椅子の識別など) では、視点変化に対して安定な識別特徴の検出が重要 (Biederman, 1985) だが、下位クラスの認識 (様々な顔の認識など) では、視点に依存した物体表現が獲得されている可能性が高い (Bülthoff & Edelman, 1992; Edelman & Bülthoff, 1992)。

一方、神経生理学の実験によれば、マカクザルの大脳皮質の下側頭野 (IT野) の前部 (TE野) には、比較的複雑なパターンに応答する細胞が存在し、類似したパターンに応答する細胞が集まってコラム (モジュール) を形成していることが示唆されている (Tanaka, 1993; Tanaka et al., 1991; Fujita et al., 1992)。特に、3次元物体に関しては、特定の視点から見た物体像に応答する細胞がIT野に存在することが確認されている (Logothetis & Pauls, 1995; Logothetis et al., 1995)。さらに、人形の顔を奥行き方向に回転しながら呈示すると、活動する領域がIT野の皮質表面を連続的に移動することが光学計測法を用いた実験で確認されており、視点に依存した表現が連続的に分布している可能性が指摘されている (Wang et al., 1996)。

次に、認識過程に関しては、網膜からV1野、V2野、V4野を経てIT野に至る側頭野への経路 (ventral pathway) が基本的な情報の流れと考えられている。Thorpeら (1996) は、心理課題 (画像に動物が存在するかを判断) を用いてヒトの脳の事象関連電位 (ERP) を測定し、判断に要する時間が極めて短い (約150msec) ことを示し、認識過程が単一方向処理であると主張している。しかし、この実験では限定された課題 (特定カテゴリーの検出) が用いられているとともに、画像が比較的単純 (対象物体は単一で存在) であるため、日常の複雑な情景では、単一方向処理だけで物体認識が可能かは疑問である。実際、上位処理系からの下降的な信号の流れを示

唆する心理学的現象として、選択的注意の現象や2値化画像の認識過程（例えばMumford, 1992）が知られている。特に、Cavanagh (1991)は、認識に係わる心理現象を分析して、陰影の解釈が入力情報だけに基づくのではなく、記憶情報が解釈に寄与していることを指摘している。さらに最近、SinhaとPoggio (1996)は、学習時に得た物体画像が「動きからの3次元構造知覚」に影響を与えることを示す心理実験を行ない、奥行き推定に記憶情報がトップダウン的に効く可能性を指摘している。

一方、生理学的には、サル脳の殆ど全ての領野間に双方向の結合が存在することが解剖学的に明らかにされている（Felleman & Van Essen, 1991）。また、電気生理学の実験からは、高次の領野からの逆投射を示唆する実験結果として、V4野の細胞の応答が入力以外の情報や注意により変化すること（Haenny et al. 1988）などが挙げられる。さらに、PETやfMRIといった非侵襲的計測装置を用いた実験により記憶からの視覚イメージの想起過程が調べられてきている。例えば、Kosslyn et al. (1995)は視覚イメージがV1野で生じていることを示唆するPETの実験結果を報告している。これは下側頭野からの逆投射によるものと考えられる。ただし、視覚イメージの生成箇所が完全に特定されるにはまだ至っていない。例えば、劣化した画像を記憶像をもとに解釈・認識するとき、下側頭野から頭頂野への投射が役立っていることを示唆するPETの報告（Dolan et al., 1997）もあり、複数の領野が同時に関与していることが考えられる。

以上の知見から、神経系における物体表現と認識過程に関して、次のようなシステムが想定できる。1) 視点に依存した物体表現がその類似性を基に自動的にクラスタリングされ、各クラスを特徴づける情報が下側頭野にコラム状に記憶される。2) 複雑な情景の物体認識においては、網膜から下側頭野に至る単一方向的な処理に限らず、複数の領野間での双方向的な情報循環が重要な役割を果たしている。ただし、コラム内の表現形式、表現獲得の仕組み、領野間で伝達される情報内容など、その詳細はまだ明かではない。計算論的神経科学の目的は、実験による知見を取り込みながら、未知の機構に対しては仮定を置いて、その実現様式を構成して試みることである。次節で述べる認識モデルは、このようなアプローチに沿って構成された一つの処理様式を示している。

4 物体認識の双方向モデル

本節で述べる3次元物体認識モデルは、混合型自己連想ネットワーク（Mixture of auto-associative networks）を用いた視点依存表現の学習と、視点推定・領域推定・画像変換補正を行ないながら物体像照合を行なう双方向緩和計算（Bidirectional relaxation）により構成されている（Ando, 1996; 藤田と安藤, 1997）。自己連想ネットワークは、恒等写像を学習する多層パーセプトロンであり、入出力層に比べて中間層のユニット数が制限された構造になっている。従って、入力像の次元を中間層で圧縮し、入力の統計的分布を実質的に規定している次元（特徴）を抽出することができる（Hertz et al. 1991; Bishop, 1995）。3層のネットワークを用いた場合、線形圧縮の一手法である主成分分析（PCA）と本質的に同等の次元圧縮を行なうことが示されている（Oja, 1991）。一方、5層（圧縮部3層と復元部3層）のネットワークを用いた場合は、非線形の次元圧縮を行なうことができるため、これまで時系列データや顔画像の圧縮（DeMers & Cottrell, 1993）、色情報の符号化（Usui et al., 1992）、物体保持のための視覚-運動情報の統合モデル（Uno et al., 1995）等に利用されてきている。さらに最近、Dayan et al. (1995)は、確率的に動作する自己連想モデルを提案し、これをHelmholtz machineと名付けている。一方、混合型の自己連想ネットワークは、複数の自己連想モジュールを用いて対象クラスごとに次元圧縮するモデルであり、類似した手法が独立に提案されている（鈴木と安藤, 1993, 1996; Suzuki & Ando, 1995; Hinton et al., 1995; Schwenk & Milgram, 1995）。特に、Suzuki & Ando (1995)は、従来の混合型回帰モデル（Jacobs et al., 1991; Jordan & Jacobs, 1994）を拡張して、教師なし学習により入力データの自動的なクラスタリングを行なう手法を提案している。

図1に示すように、本モデルの記憶部は、上記の混合型5層自己連想ネットワーク（双方向モジュール）で構成されている。入力画像は、まずガウス関数等により平滑化され、格子位置において濃淡値が抽出されて、各モジュールへの入力となる。各モジュールは、特定の物体の様々な視点方向からの投影像を圧縮し、再び入力投影像を復元するように学習が行なわれる。すなわち、この学習を通して物体像の全てがそのまま記憶されるのではなく、入力データの持つ冗長性を取り除くような情報圧縮が行われ、画像の変化を引き起こしている本質的な要因である「視点方向の情報」が抽出される。認識時には、入力の物体投影像に対して、各モジュールは視点方向を推定し、それに対応する物体像を想起して出力する。このとき、入力像を最も良好に復元するモジュールを選択し、それに対応する物体を認識物体として同定する。物体を同定する識別子（classifier）としては、以下のようにモジュールの復元誤差をsoftmax関数（Jacobs et al., 1991）：

$$f_i = \exp[-\|y^* - y_i\|^2] / \sum_j \exp[-\|y^* - y_j\|^2] \quad (1)$$

に適用したものを用いる。ここで、 y^* と y_i はそれぞれ、 i 番目のモジュールにおける入力と出力を示す。この識別子は、復元誤差の小さいモジュールに対しては1に近い値を、その他は0に近い値を出力する。

ネットワークの学習は以下に行なう。教師あり学習では、物体とモジュールの対応関係を予め指定し、入出力の自乗誤差を最小にするようにネットワークのパラメータ（シナプス荷重）を変更する。一方、教師なし学習の場合は、物体を特定せずに、複数の物体の異なる視点からの物体画像を自動的にクラスタリングしながら、各モジュールの学習を行う。この物体とモジュールの対応関係は、入力像の復元精度に応じて各モジュールの重みを変更する競合学習において獲得される。これは、目的関数：

$$\ln L = \ln \frac{\sum_i \exp[-\alpha \|y^* - y_i\|^2]}{\sum_j \exp[-\|y^* - y_j\|^2]} \quad (2)$$

の最大化により達成される（Suzuki & Ando, 1995; 鈴木と安藤, 1996）。ここで、 α はモジュール間の競合の強さを決める定数である。この最適化は少なくとも1個のモジュールの入出力を一致させる一方で、残りのモジュールの入出力誤差を増加させる作用がある。

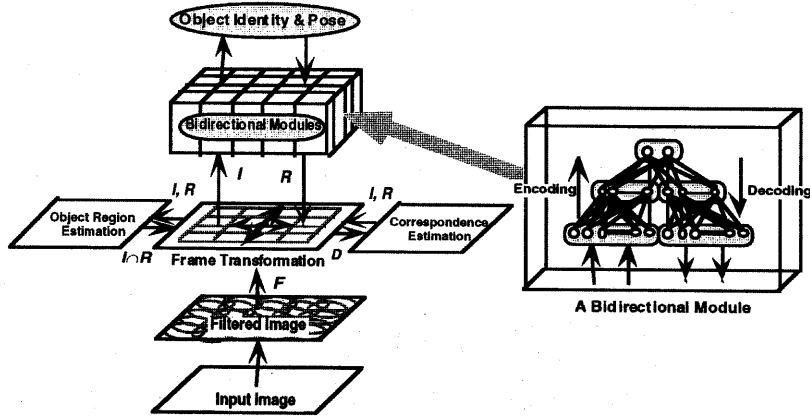


図1：物体認識の双方向モデル

次に、双方向緩和計算に基づく物体認識について述べる。一般に認識時においては、様々な視点方向からの物体像が入力されるとともに、位置変位、画像内回転、拡大縮小などの画像変換や、複数の物体の間で部分的な遮蔽が生じ得る。従って、認識時には、視点（view）の推定、物体領域の推定、変換の補正（入力像と生成像の整合）を行なうが、これらは相互依存的関係にあるため、これら全ての推定を同時に進める必要がある。よって、以下のような評価関数の最適化（コストの最小化）としてこの問題の定式化を行なう：

$$\begin{aligned} & \sum_{i,j} d_{i,j} [I(dx_{i,j}, dy_{i,j}) - R(I)]^2 \\ & + \lambda \sum_{i,j} [(dx_{i+1,j} - dx_{i,j})^2 + (dx_{i,j+1} - dx_{i,j})^2 + (dy_{i+1,j} - dy_{i,j})^2 + (dy_{i,j+1} - dy_{i,j})^2] \\ & + \eta \sum_{i,j} (1 - d_{i,j}). \end{aligned} \quad (3)$$

ここで、 I はネットワークへの入力像、 R は生成像を表わし、 $(dx_{i,j}, dy_{i,j})$ は格子点 (i, j) における変位ベクトルを表わす。 R は I 、よって $(dx_{i,j}, dy_{i,j})$ の関数である。また、 $d_{i,j}$ は物体領域か否かを示す $(1, 0)$ の2値変数（data processと呼ぶ）である。 λ と η は定数である。従って、第1項は物体領域内の入力像と生成像の整合の度合いを、第2項（正則化項）は変位ベクトルの空間的な滑らかさの度合いを、第3項は領域否定へのペナルティを表わしている。この評価関数の最適化により、変換補正を表わす $(dx_{i,j}, dy_{i,j})$ と物体領域を示す $d_{i,j}$ を求める。一方、こ

れは非線形の最適化になるため、この評価関数の平均場近似を行なうとともに、最急降下法により、「ネットワークによる視点推定・物体像生成」と「生成像と入力像の整合・物体領域の推定」を反復しながら、最適値の探索を行なう。このとき情報は、図1に示す階層的な処理系の間を双方向的に流れることになる。

5. 計算機実験

我々はこれまで、計算機で生成した物体（針金状物体）や実物体（人形）の濃淡画像を用いて、本モデルの特性を実験的に調べてきた（Suzuki & Ando, 1995; 鈴木と安藤, 1996; 藤田と安藤, 1997; Fujita & Ando, 1997）。その結果、以下のような特性が得られた。1）圧縮特性：ネットワークの学習により、視点方向に対応する情報が中間層に抽出される。2）復元特性：比較的少量の記憶表現の補間により、異なる視点からの物体像を生成できる。3）物体の識別：各ネットワークの復元精度を用いて物体の識別が可能である。4）自己想起能力：既学習のネットワークは、劣化した入力画像に対しても原画像をかなり想起できる。5）汎化特性：比較的少数の物体像の学習で、異なる視点からの物体像を生成できる。6）教師なし学習：物体のラベルを与えずに、物体像のクラスタリングが行なえる。7）変換補正と分節化：双方向緩和計算により、視点の推定、画像変換の補正、物体領域の推定が行なえる。ただし、以上述べた特性は完全なものではなく、様々な限界があることを付け加えておく。次に、この中から分節化に関連した実験を取り上げ、その結果をもう少し詳しく示しておく。

図2は、ネットワークを実物体の投影像（360度の視点変化）で学習させた後、部分的に遮蔽（欠損）した劣化画像（上段左端：平滑化画像）を入力した例を示している。ネットワークの出力画像（中段左端）は、自己想起性により、原画像に類似した画像が生成されているが、入力像は左部分が遮蔽されているため、物体の姿勢は正面から左向きにやや回転して生成されている。よって、入出力画像の非共通領域（下段左端）は遮蔽領域に必ずしも一致しない。そこで、非共通領域のみ生成画像の濃淡値で充填し、この画像を再びネットワークに入力し、同様の処理を行なう（左より2列目）。このような双方向的な処理を繰り返すことにより、徐々に遮蔽領域の推定精度が上がるとともに、視点方向（姿勢）の推定も正しい値に近づくことがわかる（左列から右列へ）。

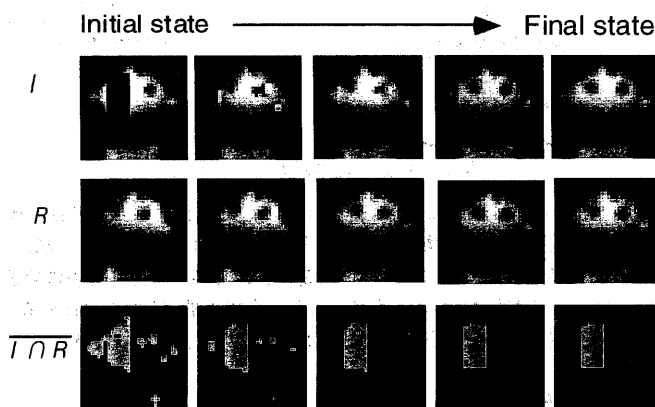


図2：双方向計算に基づく視点推定と遮蔽領域推定

図3は、複数の物体が重なって存在する情景における視点（姿勢）推定、位置補正、および分節化の実験例を示している。上段の図の太枠は、中央の物体を学習したときの注意フレームの位置を示しているのに対し、細枠はテスト時に変位した注意フレームの位置を示している。前節で述べた双方向計算の結果、変位の補正が行なわれると同時に、物体領域の推定が適切に行なわれたことが分かる（下段の図）。さらに、物体の異なる姿勢に対しても、適切に物体領域が背景領域から切り出されている。また、ここでは示されていないが、手前に位置する物体により部分的な遮蔽が生じた場合でも、視点推定・位置補正・分節化について同様な結果が得られている。

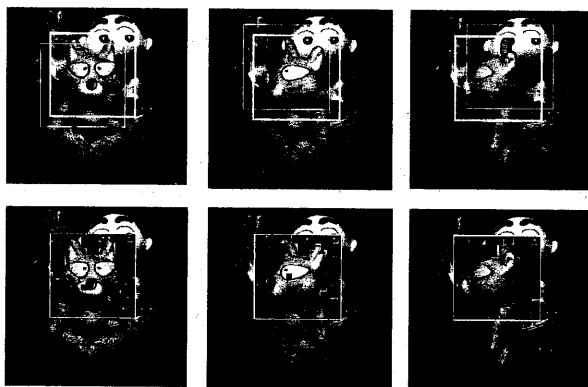


図3：複数の物体が重なって存在する情景における視点（姿勢）推定，位置補正，および分節化

6 考察とまとめ

これまでの視覚認識研究において，様々な手法や理論が提案されてきたが，全体的な柔軟性では神経システムの持つ能力にまだ遥かに及ばないと思われる．これは単に神経系における素子数の莫大さや並列処理能力といった構造的特性の違いだけでなく，基本的な計算原理が未知であるからではないだろうか．本研究では，神経系では重要な要素と考えられる「物体像表現の学習」と「双方向処理（情報循環）」を用いた一つの物体認識モデルを提示し，それが有効に機能する実験例を示した．しかし一方で，数多くの課題が将来に残されている．第一に，より複雑な画像特徴（エッジ・輪郭・多重解像度表現・3次元情報など）を用いたボトムアップ処理の充実である．第二は逆に，トップダウン情報（高次の拘束・特徴・知識・連想）の積極的な利用方法である．第三に，これらを統合するよりロバストかつ効率的なアルゴリズムの検討である．これらの可能性を様々な側面から探っていくことは，神経科学と計算機科学の両者の発展に役立つものと考ええる．

参考文献

- G. L. Agin and T. O. Binford, Computer description of curved objects. IEEE Trans. Comput., 25 (1976) 439-449
- H. Ando, 3D Object recognition using bidirectional modular networks, in: S.Z. Li, ed., Recent Developments in Computer Vision (ACCV'95), (Springer-Verlag, New York, 1996) 467-475.
- I. Biederman, Human image understanding: recent research and a theory. Computer Vision, Graphics, and Image Processing 32 (1985) 29-73.
- C.M. Bishop Neural Networks for Pattern Recognition. Oxford University Press (Oxford, UK, 1995)
- H.H. Bülthoff and S. Edelman, Psychophysical support for a two-dimensional view interpolation theory of object recognition. Proceedings of the National Academy of Science USA, 89 (1992) 60-64.
- P. Cavanagh, What's up in top-down processing?, in A. Gorea, ed., Representations of Vision. Cambridge Univ. press (1991)
- P. Dayan, G.E. Hinton, R.M. Neal, and R.S. Zemel, The Helmholtz Machine. Neural Computation, 7 (1995), 889-904
- D. DeMers and G. Cottrell, Non-linear dimensionality reduction, in: S.J. Hanson, J.D. Cowan, and C.L. Giles, eds., Advances in Neural Information Processing Systems 5, (Morgan Kaufmann Publishers, San Mateo, CA, 1993) 580-587.
- R.J. Dolan, G.R. Fink, E. Rolls, M. Booth, A. Holmes, R.S.J. Frackowiak, and K.J. Friston, How the brain learns to see objects and faces in an impoverished context. Nature, 389 (1997) 596-599.
- R.O. Duda and P.E. Hart, Pattern Classification and Scene Analysis. (John Wiley & Sons, NY, 1973).
- S. Edelman and H.H. Bülthoff, Orientation dependence in the recognition of familiar and novel views of 3D objects. Vision Research, 32 (1992) 2385-2400.
- D.J. Felleman & D. VanEssen, Distributed hierarchical processing in the primate cerebral cortex. Cereb. Cortex 1 (1991) 1-47
- I. Fujita, K. Tanaka, M. Ito, and K. Cheng, Columns for visual features of objects in monkey inferotemporal cortex. Nature, 360 (1992) 343-346.
- T. Fujita, S. Suzuki, and H. Ando. 3D object recognition by coupling mixtures of autoencoders and dynamic matching. Proc. of the Inter. Conf. on Neural Info. Proc. (Hong Kong, 1996) 377-382.
- 藤田俊史, 安藤広志 双方向処理に基づく3次元物体画像の分節化, 信学技報, NC96-46, (1996) 55-62
- 藤田俊史, 安藤広志 3次元物体表現の学習と動的認識モデル, 映像情報メディア学会誌, 51 (8) (1997), 1249-1261
- T. Fujita and H. Ando. Image segmentation for 3D object recognition using bidirectional networks. Proc. of the Inter. Conf. on Artificial Neural Networks. (1997) 943-948
- K. Funahashi, On the approximation realization of continuous mappings by neural networks. Neural Networks 2 (3) (1989)

- W.E.L. Grimson, *Object Recognition by Computer*. The MIT Press (Cambridge, MA, 1990)
- P.E. Haenny, J.H.R. Maunsell, and P.H. Schiller, State dependent activity in monkey visual cortex II. Retinal and extraretinal factors in V4. *Experimental Brain Research*, 69 (1988) 245-259.
- J. Hertz, A. Krogh, & R.G. Palmer, *Introduction to the Theory of Neural Computation*. Addison Wesley (Redwood City, CA 1991)
- G.E. Hinton, M. Revow, and P. Dayan, Recognizing handwritten digits using mixtures of linear models, in: G. Tesauro, D. Touretzky, and T. Leen, eds., *Advances in Neural Information Processing Systems 7*, (The MIT Press, Cambridge, MA, 1995) 1015-1022.
- R.A. Jacobs and M.I. Jordan, Learning piecewise control strategies in a modular neural network architecture. *IEEE Transactions on Systems, Man, and Cybernetics*, 23 (2) (1993) 337-345.
- R.A. Jacobs, M.I. Jordan, S.J. Nowlan, and G.E. Hinton, Adaptive mixtures of local experts. *Neural Comp.*, 3 (1991) 79-87.
- M.I. Jordan and R.A. Jacobs, Hierarchical mixtures of experts and the EM algorithm. *Neural Comp.*, 6 (2) (1994) 181-214.
- M. Kawato, H. Hayakawa, and T. Inui, A forward-inverse optics model of reciprocal connections between visual cortical areas. *Network* 4 (1993) 415-422.
- S.M. Kosslyn, W.L. Thompson, I.J. Kim, and N.M. Alpert, Topographical representations of mental images in primary visual cortex. *Nature* 378 (1995) 496-498.
- N.K. Logothetis and J. Pauls, Psychophysical and physiological evidence for viewer-centered object representations in the primate. *Cerebral Cortex*, 3 (1995) 270-288.
- N.K. Logothetis, J. Pauls, and T. Poggio, Spatial reference frames for object recognition tuning for rotations in depth., A.I. Memo No. 1533, C.B.C.L. Paper No. 120, MIT, 1995.
- D.G. Lowe, Three-dimensional object recognition from single two-dimensional images. *Artif. Intelligence*, 31 (1987) 355-395.
- D. Marr and H.K. Nishihara, Representation and recognition of the spatial organization of three-dimensional shapes. *Proc. Royal Society of London B*, 200 (1978) 269-294.
- D. Mumford, On the computational architecture of the neocortex II. *Biol. Cybern.* 66 (1992) 241-251.
- J. L. Mundy and A. Zisserman eds., *Geometric Invariance in Computer Vision*. MIT Press, Cambridge, MA (1992)
- H. Murase and S.K. Nayar, Visual learning and recognition of 3D objects from appearance. *International Journal of Computer Vision*, 14 (1) (1995) 5-24.
- S. K. Nayar and T. Poggio eds., *Early Visual Learning*. (Oxford University Press, Oxford, UK, 1996)
- E. Oja, Data compression, feature extraction, and autoassociation in feedforward neural networks, in: T. Kohonen, et al. eds., *Artificial Neural Networks*, (Elsevier Science publishers B.V., North-Holland, 1991) 737-745.
- B.A. Olshausen, C.H. Andersen, and D.C. Van Essen, A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *The Journal of Neuroscience*, 13 (11) (1993) 4700-4719
- T. Poggio and S. Edelman, A network that learns to recognize three-dimensional objects. *Nature*, 343 (1990) 263-266.
- R.P.N. Rao and D.H. Ballard, Dynamic model of visual recognition predicts neural response properties in the visual cortex. *Neural Computation*, 9 (1997) 721-763.
- H. Schwenk and M. Milgram, Transformation invariant autoassociation with application to handwritten character recognition, in: G. Tesauro, D. Touretzky, and T. Leen, eds., *Advances in Neural Information Processing Systems 7*, (The MIT Press, Cambridge, MA, 1995) 991-998.
- P. Sinha and T. Poggio, Role of learning in three-dimensional form perception. *Nature* 384 (1996) 460-463
- 鈴木敏, 安藤広志, モジュール学習による3次元物体の認識と類別. 信学技報, NC93-62, (1993)59-66
- 鈴木敏, 安藤広志, 2次元射影像からの3次元物体の認識と類別 --モジュール構造を用いた教師なし学習モデル--.
- 電子情報通信学会論文誌, J79-D-II, 7, (1996)1291-1300
- S. Suzuki and H. Ando, Unsupervised classification of 3D objects from 2D views, in: G. Tesauro, D. Touretzky, and T. Leen, eds., *Advances in Neural Information Processing Systems 7*, (The MIT Press, Cambridge, MA, 1995) 949-956.
- K. Tanaka, Neuronal mechanisms of object recognition. *Science*, 262 (1993) 685-688.
- K. Tanaka, H. Saito, Y. Fukada, and M. Morita, Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *Journal of Neurophysiology*, 6 (1) (1991) 170-189.
- S. Thorpe, D. Fize, and C. Marlot, Speed of processing in the human visual system. *Nature*, 381 (6) (1996) 520-522.
- S. Ullman and R. Basri, Recognition by linear combinations of models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13 (10) (1991) 992-1006.
- S. Ullman, *High-level Vision: Object Recognition and Visual Cognition*, The MIT Press (Cambridge, MA, 1996)
- Y. Uno, N. Fukumura, R. Suzuki, and M. Kawato, A computational model for recognizing objects and planning hand shapes in grasping movements. *Neural Networks*, 8 (6) (1995) 839-851.
- S. Usui, S. Nakauchi, and M. Nakano, Reconstruction of Munsell color space by a five-layer neural network. *Journal of Optical Society of America A*, 9 (4) (1992) 516-520.
- G. Wang, K. Tanaka, and M. Tanifuji, Optical imaging of functional organization in the monkey inferotemporal cortex. *Science*, 272 (1996) 1665-1668.
- L. Wiskott and C. von der Malsburg, A neural system for the recognition of partially occluded objects in cluttered scenes: a pilot study. *Int. J. of Pattern Recognition and Artificial Intelligence*, 7 (4) (1993)