

視覚の計算論的モデル テクスチャーと遮蔽と3次元知覚

酒井 宏

理化学研究所 国際フロンティアシステム 脳回路モデル研究チーム
〒351-01 埼玉県和光市広沢2-1

1. 緒言

近代哲学の基礎を固めたとされるデカルト (Descartes) は精神と身体は完全に独立していると説いたが、今日では心理的なプロセスも脳内の物理化学的反応によって引き起こされると考えられるようになった。脳は神経細胞の集まりに他ならないが、単一の細胞自体では行い得ない思考や感情、記憶や認知といった心理プロセスを実現していると考えられている。高次の視覚機能も脳内の多数の神経細胞による回路網によって実現されている。近年の生理学、心理物理学の発展と、計算機、シュミレーション技術の進歩によって、視覚機能がどのような論理や回路網によって実現されているのかを探る研究が可能になってきた。ここでは、それらの回路網とそこで実行されているアルゴリズムを解明する事を目的として構築された、最近の計算論的モデルについて概説する。

2. 3次元空間知覚のリバースエンジニアリング

脳は神経細胞の集まりに他ならないが、単一の細胞自体では行うことのできない知覚や認知といった高次プロセスを実現している。高次視覚機能も実際にはイオンの移動、膜電位の変化といったミクロな現象によって実現されている訳であるが、高次機能のプロセスで一つ一つの細胞の役割や機能を実験的に解明することは困難である。また、現在の計算機には微細な構造を加味して高次機能をシュミレーションする容

量と能力はない。中高次視覚機能を理解していく上では、独立した細胞一つ一つの機能を考えることを離れ、多くの細胞が集団 (population) として機能することを念頭に置くことが必要である。

知ろうとする機能が高次になるにつれて、生理学的実験によって知見を得ることが一般に難しくなる。中高次機能では、心理物理学の知見や計算理論が重要な証拠を提供する。しかし、機能が神経細胞の集団によって容易に実現されるだろうことは依然として大切である。このように、中高次機能では生理、解剖学、心理物理学、そして計算理論を統合してメカニズムを解明していくことが重要である。脳をブラックボックスとして、知ることのできる情報からその内部メカニズムを推測することから、この手法はリバースエンジニアリングであると言えよう。

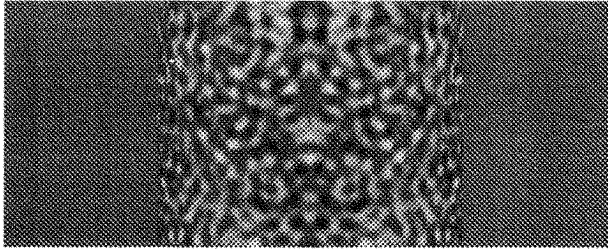
ここでは、重要な視覚機能の一つである、3次元知覚に関する2つのモデルを概説する。我々は3次元の世界に住んでいるが、その画像はいったん2次元の網膜に投影される。我々が内部に知覚する3次元構造は、脳がその2次元の画像から再構築したものなのである。3次元から2次元への投影の過程において、多数の物体が同じ画像に投影されてしまう。このために、2次元画像から3次元構造の再構成は難しい問題となる。視覚系がどのようにこの問題を解いているかは視覚研究の主要なトピックの一つといえよう。3次元知覚では両眼の視差によ

るステレオ視があるが、写真から容易に3次元構造が知覚できることから明らかなように、単眼から得られる3次元情報も重要である。陰影、テクスチャーの変化、後ろの物体が前の物体に隠される遮蔽 (occlusion)、などは強力な3次元知覚を引き起こす。ここでは、テクスチャーから3次元形状を知覚するモデルと、遮

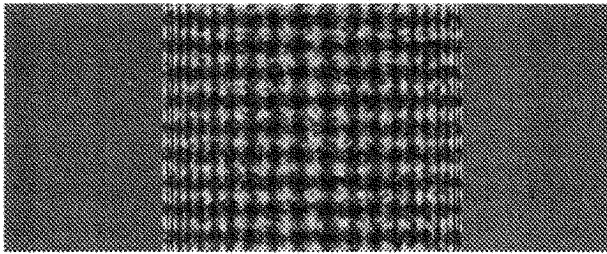
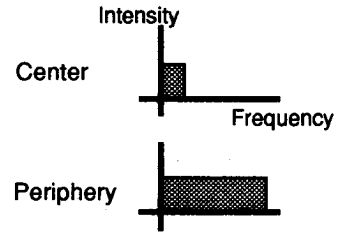
蔽から3次元構造を知覚するモデルを紹介する。

3. テクスチャー変化からの3次元形状知覚モデル

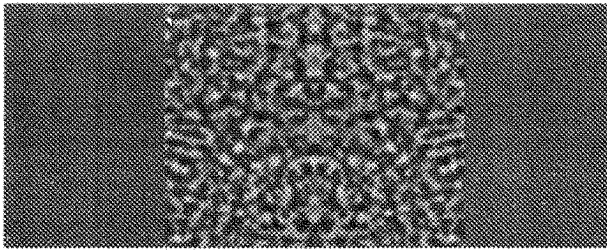
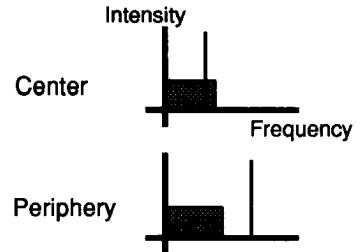
テクスチャーをもつ表面が傾斜すると、傾斜に従ってテクスチャーが変化して見える。テク



(A)



(B)



(C)

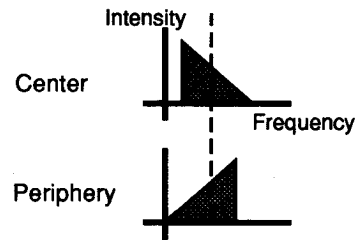


図1 ホワイトノイズを周波数領域でフィルタリングすることによって作った、人工的な周波数スペクトラムをもつ刺激の例。左側にグレーを背景とした3つの刺激と、右側にそれらの周波数スペクトラムの水平方向成分を示した。スペクトラムの特徴によって3次元的な円柱が知覚されたり、されなかったりする。(A) スペクトラムは、中心領域では低周波のホワイトノイズ成分からなり、左右に移動するに連れて順次高周波成分まで含むように変化する。生成された刺激は垂直に置かれた円柱に見える。(B) 刺激の全領域に、同一の低中周波ホワイトノイズ成分がある。これに加えて強い単一のピークがあり、これは中心から左右の周辺領域に向かって移動するに連れて、低周波から高周波へと変化する。ピークは単一成分なので、このスペクトラムの平均周波数はホワイトノイズ成分によって決まり、この刺激の場合は全域で同一である。このように強いピークが変化する場合には、平均周波数に変化がなくとも、生成された刺激はやはり3次元的に見える。(C) スペクトラムにピークが存在するが、その強度が比較的弱い場合。平均周波数は全領域で一定になるように制御されている。ピークが弱い場合はそれが適当に変化しても、生成された刺激は3次元的には見えない。詳細は Sakai and Finkel²⁾ を参照。

スチャーはその Fourier スペクトラムによって一意に記述できることから、理論的にはスペクトラムの変化から傾斜が計算できる。コンピュータビジョンのアルゴリズムとしては、2か所のスペクトラムの間のアフィン変換を求めることによって相対的な傾斜を求めるアルゴリズムが提案されている（例えば Malik and Rosenholtz¹⁾）。しかし現実には、複雑な2次元スペクトラム間の変換を求める計算は容易でなく、こういった計算が視覚系で行われているとは考えにくい。

視覚系では、複雑なパラメーターから知覚を行う場合に特徴化をおこなう例が知られている。例えば色の認識では、連続したスペクトラムそのものではなく、3つの波長によって代表される特徴量が使われていると考えることがで

きる。テクスチャーから傾斜を知覚する場合にも、このような特徴量が使われていることを示す心理物理実験が報告されている（Sakai and Finkel²⁾）。実験では、ホワイトノイズを Fourier 領域でフィルタリングすることによって任意の特徴を持つ人工的なスペクトラムからならさまざまな刺激を作った。そしてこれらの刺激を被験者に提示することによって、その特徴がヒトの3次元知覚に及ぼす影響を求めた。図1に刺激の例を幾つか示す。実験の結果から、視覚系ではスカラー量である少数の平均ピーク周波数（average peak frequency）によって傾斜を計算していることが示唆された。

平均ピーク周波数は、局所のピーク周波数を空間的に平均化して求められる量で、Fourier 領域では次のように表される。

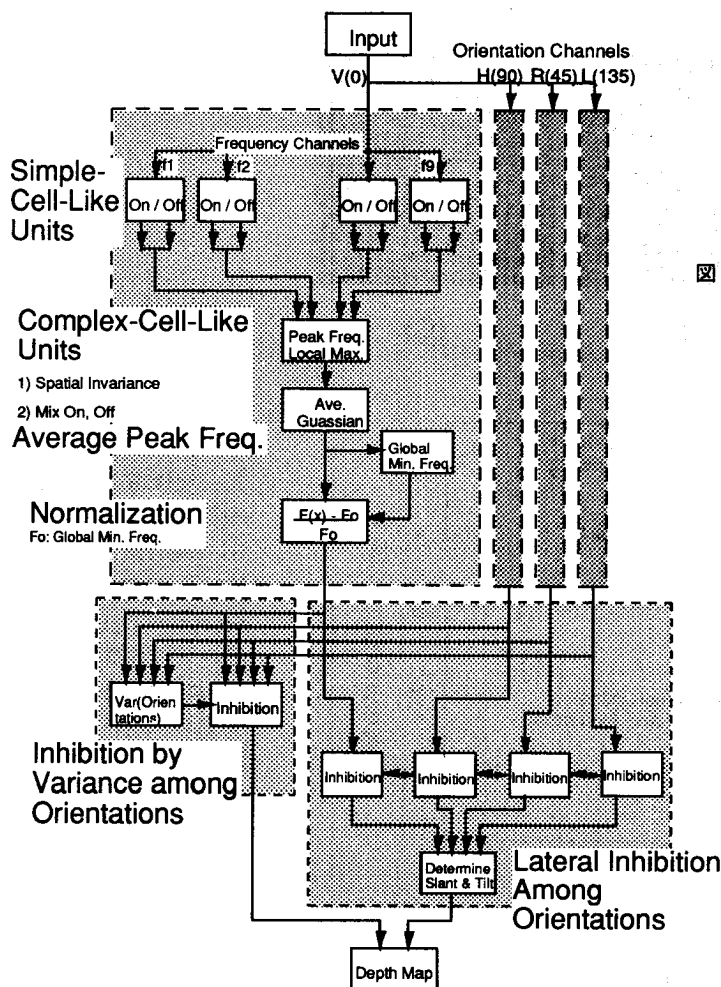


図2 テクスチャー変化から3次元形状を知覚するネットワークモデル。入力画像はV1単純型細胞と同様な受容野によってV,H,R,Lの方向性チャンネルに分解される。さらにその出力は複雑型細胞のモデルによって空間的にブールされ、テクスチャーに相当する周波数成分が抽出される。抽出された周波数情報は、平均ピーク周波数（APF）によって特徴化される。APFは各方向ごとに正規化され、その方向への傾斜の程度を示す。その後、側方結合によって方向性チャンネル間の融合が行われ、3次元形状の情報が得られる。

$$\bar{f}_p^0 = \int_R |u|N(u,v)du dv \quad (1)$$

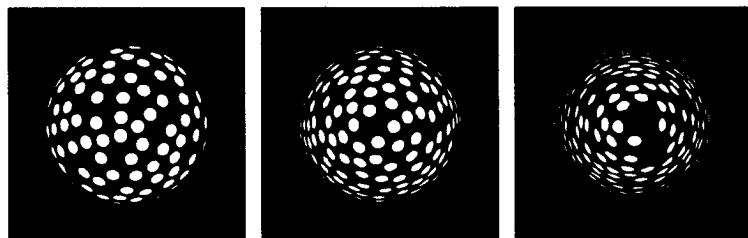
これは水平方向 ($\phi=0$) に沿った平均ピーク周波数で、 R は水平方向の狭い扇形領域を示す。 $N(u,v)$ は、この成分がピークとして選ばれる回数である。平均ピーク周波数は、各周波数成分がそれぞれ最大のエネルギーをもつ確率を示す確率密度関数の性質によって、平均周波数とピーク周波数の間の値を取る。たとえば、確率密度関数がスペクトラム分布と一致する場合は、平均ピーク周波数は平均周波数と一致する。

Sakai and Finkel³⁾は、平均ピーク周波数による空間周波数スペクトラムの特徴化とそれに続く正規化によって、テクスチャーの変化から3次元形状を導くネットワークモデルを提案した。図2にモデルの模式図を示す。特徴量であ

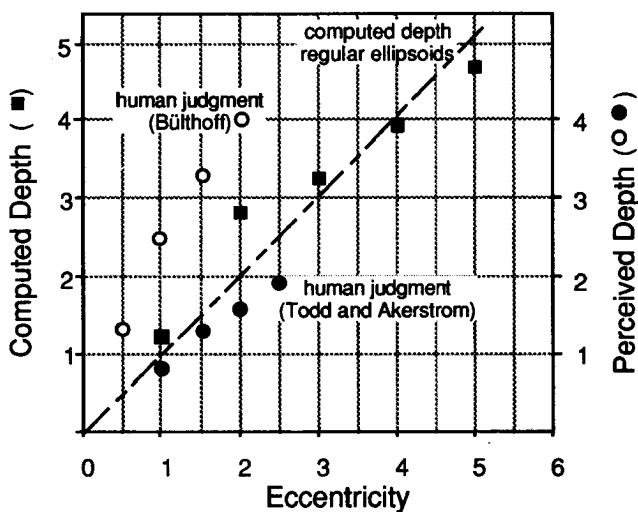
る平均ピーク周波数を正規化することによって各方位に沿った相対的な傾斜を算出し、それから側方抑制のメカニズムによって最大傾斜 (slant) とその方向 (tilt) を求めている。このモデルの反応を評価するために、離心率が異なる楕円体などについてシュミレーションをおこなった。その結果は類似の心理物理実験 (Todd and Akerstrom⁴⁾; Bülthoff⁵⁾) と比較され、両者が良く一致することが示された。用いた刺激と、離心率と反応の関係を、図3に示す。また、実画像から算出された3次元形状の例を図4に示す。複雑なテクスチャーをもつ実画像からも適当な3次元形状が求められている。

4. 遮蔽からの3次元奥行き知覚モデル

アリストテレス (Aristotle) は、視覚のタスクとは「何処に、何があるか」を知ることであ



(A) (B) (C)



(D)

図3 様々な離心率をもつ楕円体を与えた時にネットワークが計算した楕円体周辺部の奥行き。(A)もちいた離心率1の楕円体(球)。長軸方向へのorthographic projection。(B)同様の離心率2の楕円体。(C)離心率4の楕円体。(D)ネットワークが計算した奥行き(■)と、同様の心理物理学的実験の結果(●, ○)。モデルの出力は、離心率に対してほぼ線形になっており、ヒトの知覚とよく一致している。

ると定義した。もしも読者がこの本を図書館で読んでいるのなら、網膜には本と、机と、たぶん背景に本棚と居眠りをしている他の来館者が

投影されているだろう。しかし網膜そのものには、どれが本で、どの部分が机なのかは判っていない。感覚の段階では「物体」は定義されな

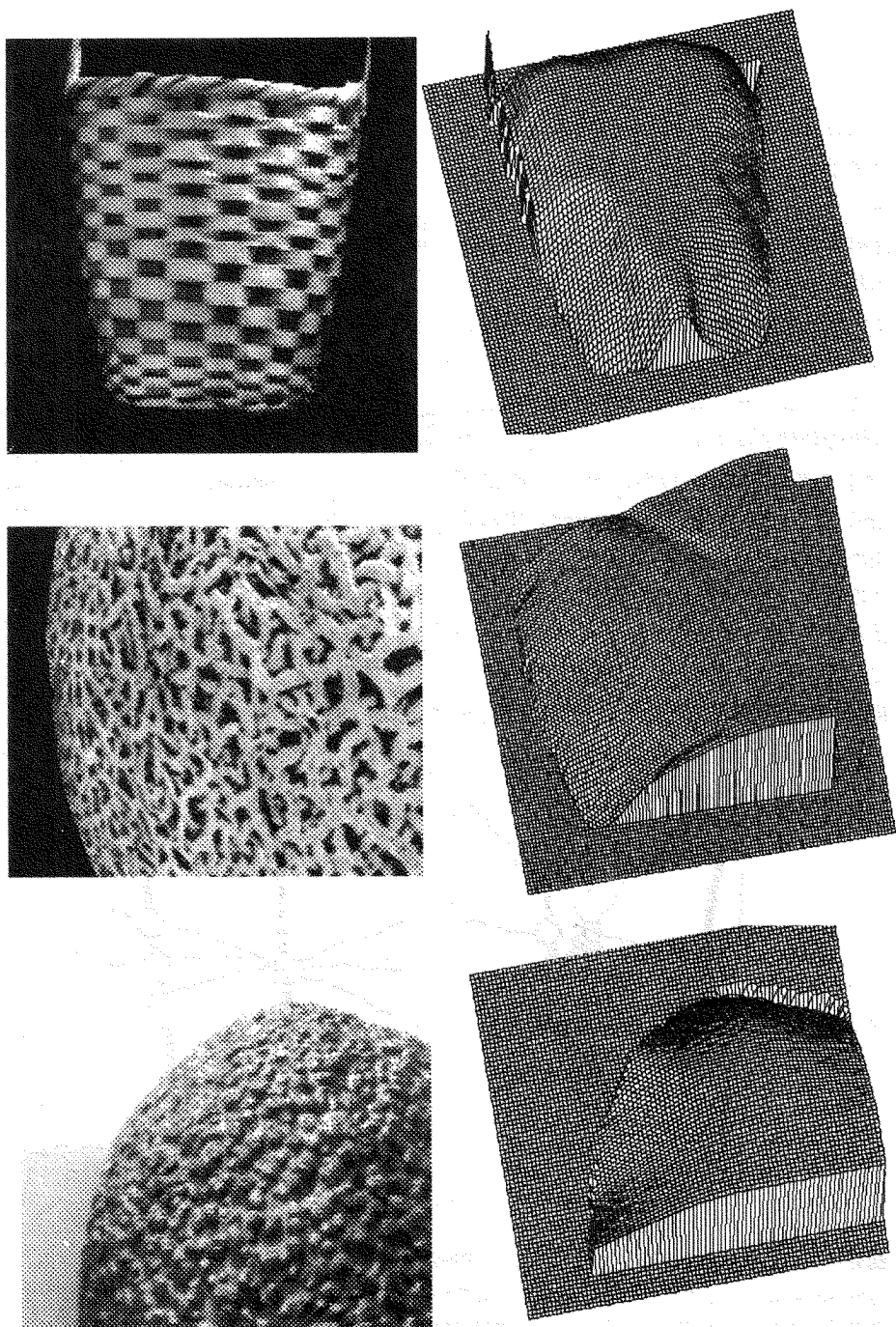


図4 ネットワークモデルがビデオ画像から算出した3次元形状。左側が入力画像、右側が出力された奥行き。上段はバスケット、中段はメロン (cantaloupe) の一部、下段はアボガドの一部。複雑なテクスチャーからでも、3次元形状を良く再現している。

いのである。これが感覚 (sensation) と知覚 (perception) の根本的な違いである。知覚過程で、画像は部分に分割 (segmentation) され、そして必要があれば幾つかの部分が結合 (binding) されて、物体が定義されると考えられる。このように知覚は、画像のどの特徴によって情景を分割し、群化すべきかを決定するプロセスを含んでいるはずである (Finkel and Sajda⁹⁾。

物体が何処にあるかを決定することは、画像中のどの部分が物体の表面であるかを決定することである。ある特徴量、多くの場合濃淡が不連続になっている部分が物体の輪郭である可能性が高い。ここで表面を定義するためには、輪郭がどちら側の領域の端を示しているのかを決めなければならない。すなわち、どちら側が図 (figure) で、どちら側が地 (ground) であるかを決めなければならない。机の上に本が置かれている場合、本の周囲に輪郭が見られるが、これは本の領域の端であり、本の輪郭なのであって、机の輪郭ではない。このような場合、本が輪郭を所有しているという。そして輪郭を

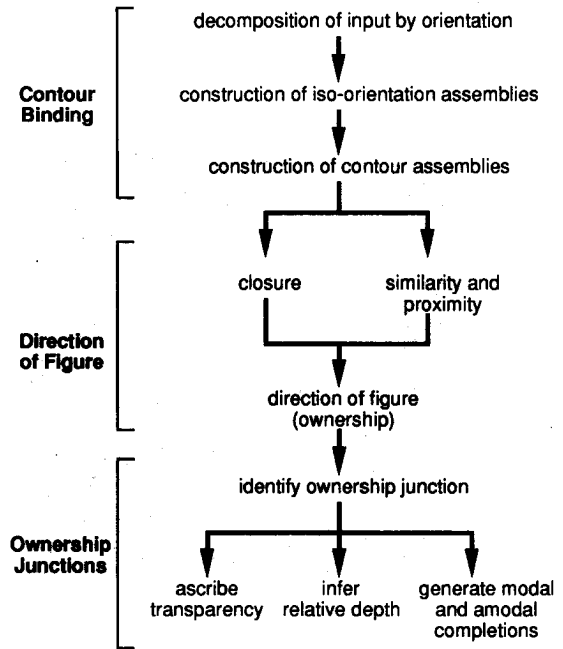
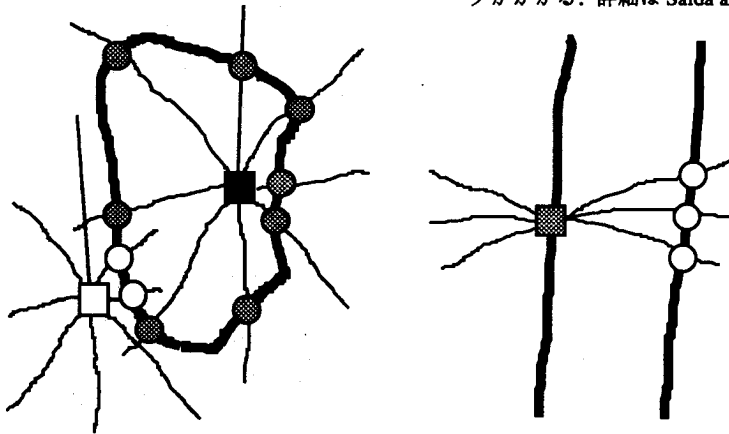


図5 遮蔽から奥行きを知覚するネットワークモデル。大きく分けて、輪郭抽出と結合、表面分割 (direction of figure)、奥行き知覚 (ownership junction) の3つの部分からなる。ここでは1次的な流れだけを示したが、実際には相互に結合されており、2サイクル目からはフィードバックがかかる。詳細は Saida and Finkel⁹⁾ を参照。



Closure

Similarity/ Proximity

図6 輪郭がそのどちら側の表面に属しているか (Direction of Figure) を決める2つのメカニズム。左側は閉合性 (closure) を検査するメカニズムを示している。四角 (■, □) が検査をするユニットで、8方に突起 (dendrite) をのばしている。1つの突起が輪郭に接触していると、ユニットの活動度が1だけ上がる。輪郭に囲まれた■の活動度は高くなるが、外側の□は活動度が低くなり、その差から、輪郭のどちら側が閉じているかが判断できる。右側は同等性、近接性 (similarity, proximity) を検査するメカニズムを示している。四角が検査をするユニットで、左右両方向に突起をのばしており、輪郭と同じ方向性 (同等性) をもっている他の輪郭が近く (近接性) にあると、活動度が上がる。

所有する側である本が図, 所有しない側である机が地となるのである. Nakayama and Shimojo⁷⁾はこれを「ownership」を決めるプロセスと呼び, 相対的な奥行きを決定する役割を果たしていることを示した.

Sajda and Finkel⁸⁾は, ownership に立脚したネットワークモデルを構築して, 表面を知覚する視覚過程を研究した. モデルは大きく分けて3つの部分からなっている. 輪郭を結合するプロセス, ownership によって図の方向 (direction of figure) を決めるプロセス, そして ownership の分岐点を基にして3次元構造を求めるプロセスである. これらのプロセスは互いに双方向に結合されている. 図5にモデルの模式図を示す. 構築されたモデルの主要部分は, 群化に関するゲシュタルト法則 (Gestalt law) をネットワークメカニズムとして実現した形になっている.

モデルの初段部分は単純型, 複雑型細胞の機能を模したユニットからなっており, これらが入力を方位情報に分解する. これらの出力は輪郭結合のプロセスに送られ, ここで類似の方向性をもつユニット同士が結合されて輪郭の線分が決められる. これは, 視覚野では類似の方位選択性をもつ細胞同士が強く結合していることに対応し, Finkel らはこの回路がゲシュタルト法則の「よい連続」 (good continuation) を実現しているのだらうと示唆している. さらに, 線分終端部分で異なる方向性をもつ他の線分と結合されて, 輪郭全体が求められる. この時, 結合は同時に発火することによって表現される. このような時間的なコーディングが脳内にあることは Gray and Singer⁹⁾らによって示唆されている.

求められた輪郭は次に表面に結合 (bind) される. 結合は輪郭同士の距離, 相対的な方位,

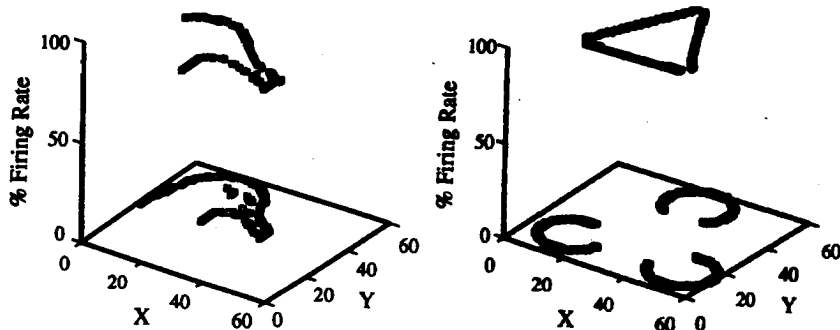
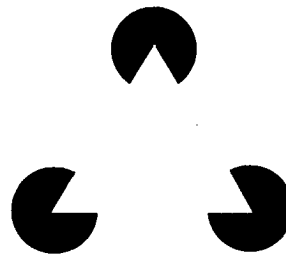
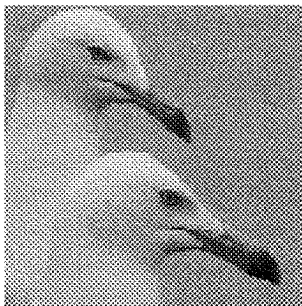


図7 入力とネットワークの反応. 上段が入力, 下段がネットワークが計算した相対的な奥行き. 活動度が低いほど遠くを示す. 左側の図では, 2羽が分割され, 下側のカモメが近くにいることが知覚されている. 右側は Kanizsa の三角形で, ネットワークが3角形の主観的輪郭を知覚し, それが3つの円より手前にあることが判る.

文 献

輪郭が閉じているかどうかによって決められる。輪郭同士は、その距離が近いほど、方位が似ているほど、同じ表面の輪郭であると判断されやすい。同時に、輪郭が閉じていれば、その内側が表面である判断されやすい。これらはゲシュタルト法則では、同等性、近接性、閉合性 (similarity, proximity, closure) と呼ばれているものである。これらのメカニズムは図6に示すような簡単なネットワークによって構成され、その出力は ownership を示すことになる。この表面知覚のメカニズムは、コントラスト情報に限らず、テクスチャ境界をもとにした表面分割にも適応できる (Sakai, Sajda, and Finkel¹⁰⁾。

輪郭の ownership が変化することは異なる物体の存在を示唆するから、ownership の分岐点から遮蔽の境界を検出できる。ここで重要なのは、ownership 分岐点は表面の境界を示すものであって、輪郭の交差を示す、いわゆる T 型分岐点とは違う意味を持つことである。透明視は濃淡情報だけからでは同定できない事が知られているが、ownership から濃淡境界部での図と地の関係を与えると透明であることが判定できる。図7に幾つかの入力に対して相対的な奥行きを求めたシュミレーションの結果を示す。遮蔽、錯視的輪郭のいずれの場合もヒトと同様な反応を示している。

5. 結言

ごく簡単ではあるが、視覚の計算論的モデルの例として、テクスチャーから3次元形状を知覚するモデルと、遮蔽を手掛かりとした表面分割と奥行き知覚に関するモデルについて述べた。V1細胞からゲシュタルト法則まで、広範囲の知見を統合する事により、従来見られなかった、より高次の視覚機能のモデルが構築できたと言えよう。

- 1) J. Malik and R. Rosenholtz: A differential method for computing local shape-from-texture for planar and curved surface. *Proceedings of CVPR 1993*, 267-273, 1993.
- 2) K. Sakai and L. H. Finkel: Characterization of the spatial-frequency spectrum in the perception of shape from texture. *Journal of the Optical Society of America A*, 12, 1208-1224, 1995.
- 3) K. Saka and L. H. Finkel: A network mechanism for the determination of shape-from-texture. *Advances in Neural Information Processing System 6*, Morgan Kaufmann Pub, San Fransisco, 953-960, 1994.
- 4) J. T. Todd and R. A. Akerstrom: Perception of three-dimensional form from patterns of optical texture. *Journal of Experimental Psychology*, 13, 242-255, 1987.
- 5) H. H. Bülthoff: Shape from X: Psychophysics and computation. *M. S. Landy and J. A. Movshon (eds): Computational Models of Visual Processing*, MIT Press, Cambridge, 1991.
- 6) L. H. Finkel and P. Sajda: Constructing visual perception. *American Scientist*, 82, 224-237, 1994.
- 7) K. Nakayama and S. Shimojo: Toward a neural understanding of visual surface representation. *Cold Spring Harbor Symposia on Quantitative Biology*, 55, 911-924, 1990.
- 8) P. Sajda and L. H. Finke: Intermediate-level visual representations and the construction of surface perception. *Journal of Cognitive Neuroscience*, 7, 267-291, 1995.
- 9) C. M. Gray and W. Singer: Neuronal oscillations in orientation columns of cat visual cortex. *Proceedings of National Academy of Science USA*, 86, 1698-1702, 1989.
- 10) K. Sakai, P. Sajda and L. H. Finkel: Texture discrimination and binding by a modified energy model. *Proceedings of International Joint Conference on Neural Networks*, vol. III, 789-795, 1992.