

**■S3 群 (脳・知能・人間) - 2 編 (感覚・知覚・認知の基礎)****8 章 3D 物体の認知**

(執筆者：笹岡貴史) [2008 年 5 月 受領]

**■概要■**

三次元物体の景観 (view) は様々な要因によって無限に変化しうる。しかし、それにもかかわらず我々は一貫した物体の認知が可能である。このような物体認知の特性を説明するために、三次元構造ベースの脳内表現と二次元画像ベースの脳内表現という二つの物体表現理論が提案され、激しい議論が行われてきた。一方、神経心理学、電気生理学的研究、更に、近年の脳機能イメージング技術の発達により、物体認知にかかわる神経基盤について徐々に明らかになりつつある。今後、心理学研究、脳研究が相補的に進展することで 3D 物体認知を説明する統一的なモデルの構築が期待される。

**【本章の構成】**

本章では、まず、物体恒常性 (8-1 節) で三次元物体認知研究において最も重要なテーマである「物体恒常性」について述べ、物体表現理論 (8-2 節) では、物体恒常性を説明するために提案されている二つの物体表現理論について紹介する。三次元物体認知にかかわる神経基盤 (8-3 節) では、電気生理学的研究、及び近年の脳機能イメージング研究によって明らかにされている三次元物体認知にかかわる神経基盤について述べ、最後に、今後の展望 (8-4 節) で三次元物体認知研究の今後の展開について述べる。

## ■S3 群 - 2 編 - 8 章

### 8-1 物体恒常性

(執筆者：笹岡貴史) [2008年5月 受領]

三次元物体の景観 (view) は照明の変化や、物体までの距離の変化、また視点の変化など様々な要因によって無限に変化する。しかし、それにもかかわらず我々は一貫した物体の認知が可能である。このことを物体恒常性 (object constancy) と呼ぶ。三次元物体認知研究の最大のテーマはこの物体恒常性がどのようにして達成されているのかを解明することであるといっても過言ではない。

#### 8-1-1 大きさ・位置不変性

まず、物体の網膜像を変化させる要因として、並進 (translation) や、物体までの距離があげられる。例えば、並進によって物体の網膜像の位置は変化し、物体までの距離が変化することで物体の網膜像の大きさは変化する。このような変化に対する不変性を大きさ不変性 (size invariance) ・位置不変性 (translation invariance) と呼ぶ。これまで心理学的研究によって、これらの不変性について実験的に示されている (例えば 2), 11), 15)。しかし、ランダムドットパターンのような新奇な図形を刺激として用いた場合に不変性が成り立たないという報告もある<sup>8)</sup>。

#### 8-1-2 視点不変性

もう一つの要因として視点の変化があげられる。視点の変化に対する物体認知の不変性は視点不変性 (viewpoint invariance) と呼ばれる。この視点の変化として  $x$  軸回転 (ピッチ),  $y$  軸回転 (ヨー),  $z$  軸回転 (ロール) が考えられるが、特に、 $x$  軸,  $y$  軸回転による視点の変化によって、視点が変化する前に見えていた視覚的特徴が遮蔽 (occlusion) される、または、逆にそれまで見えていなかった特徴が現れる場合がある。よって、このような  $x$  軸,  $y$  軸回転に対する視点不変性は、大きさや位置不変性と異なり、奥行き情報を有する三次元物体特有の現象であるといえる。

## ■S3 群 - 2 編 - 8 章

### 8-2 物体表現理論

(執筆著：笹岡貴史) [2008年5月 受領]

#### 8-2-1 三次元構造ベースの物体表現

物体恒常性を達成するためにはどのようにして物体が表現されている必要があるだろう？ この問題を説明するために、大別して2種類の物体表現理論が提案されてきた。一つは、三次元構造ベースの物体表現である。その先駆となったのが Marr と Nishihara による物体表現理論である<sup>23)</sup>。彼らの理論では、三次元物体は任意の二次元形状の断面を軸に沿って移動させた軌跡として定義される一般化円錐 (generalized cone) の組み合わせによって、物体中心座標系 (object-centered coordinate system) に基づいて表現される。このように観察者の視点と独立な物体中心座標系によって物体が表現されていれば、入力画像の視点にかかわらず同じ物体表現が復元できると考えられる。また、Biederman が提唱した GSD 理論<sup>3)</sup>では、ジオン (geon: geometric ion) と呼ばれる単純な三次元形状のセットが定義されており、その組み合わせとジオンどうしの位置関係の記述 (構造記述: structural description) によって物体が表現される。ジオンはそれ自身が視点不変的な特徴をもつ物体中心座標系で記述されており、その組み合わせによって物体が表現されていれば、視点不変な物体認知を達成できると考えられる。Biederman らのグループはジオンで構成された新奇物体を用いて数多くの心理実験を行い、この理論の正当性を主張した (例えば 4, 5)。また、線画から構造記述表現を抽出するニューラルネットワークモデルも提案されている<sup>20)</sup>。しかし、物体を構成するパーツが多くなれば組み合わせ爆発が起こることや、自然画像からこれらのジオンを抽出する方法が確立されていないといった計算論的な問題点も指摘されている<sup>9)</sup>。

#### 8-2-2 二次元画像ベースの物体表現

三次元構造ベースの物体表現理論は、三次元物体の認知における視点不変性を説明する理論として提案されたものといえる。しかし、1980年代後半から三次元物体の認知の視点依存性を示す報告が増えてくる。例えば、Rock と DiVita は針金を曲げたような新奇な三次元物体を用いて実験を行い、このような物体に回転を加えることで同じ物体であると認識することが困難になることを示した<sup>25)</sup>。また、Bülthoff と Edelman<sup>7)</sup>は、ペーパークリップオブジェクトと呼ばれる針金を折り曲げたような新奇物体を 3D コンピュータグラフィックスで作成し、それを刺激とした心理実験を行った。実験協力者は水平回転で互いに 75° 離れた二つの視点からの景観を学習し、その後、学習した景観を内挿する景観や、外挿する景観、垂直回転させた景観、または異なる物体の景観が呈示され、実験協力者はそれらが学習した物体と同一かどうか判断を行った。その結果、実験協力者の認識パフォーマンスは記憶された景観の間を補完するような視点依存性を示した。以上のような知見は、二次元的な景観に基づく (view-based) 物体表現が行われていることを示唆している。

二次元画像ベースの物体表現に基づく数理モデルもこれまで多数提案されている。例えば、Ullman は、ごく少数の景観を保持していればそれらの線形計算によって任意の view を生成可能であることを示した<sup>31)</sup>。また Poggio と Edelman<sup>26)</sup>は、GRBF ネットワークと呼ばれる三層のネットワークモデルにペーパークリップオブジェクトの頂点座標や関節角を学習させる

ことで、ペーパークリップオブジェクトのどの方位からの景観をネットワークに入力しても、あらかじめ定められた一つの景観 (standard view) を出力することが可能になることを示した。更に、ある景観が入力されたときには、その standard view からの回転角度も出力される。この Poggio と Edelman のモデルの中間層で学習される内部表現は、まさに学習した物体に関する景観ベースの物体表現と考えることができる。

しかし、二次元画像ベースの物体表現理論では、構造を明示的に表現しないために次のような問題も指摘されている<sup>21)</sup>。例えば、我々は、球の上に立方体に乗っているような物体と、立方体の上に球に乗っている物体の違いを「立方体と球の位置関係の違い」として認識できるが、景観ベースの物体表現理論では全体的な画像としての違いでしか表現できない。すなわち、二次元画像ベースの物体表現理論では、このような単純な位置関係の違いを明示的に表現することができない。

### 8-2-3 物体表現理論についての論争

両理論において議論の中心となる問題は、既知の景観から回転を加えられた未知の景観を認識するためにかかる回転コストである。これは、例えば反応時間や誤答率といった指標で測ることができる。例えば、二次元画像ベースの物体表現理論では、既知の景観と未知の景観の間の回転角度差が大きいほど同じ物体であると認識できるまでの反応時間がかかり、誤答率が高くなることが予想される。例えば、Tarr はこのような現象を「複数の景観+変換 (multiple-views-plus-transformation)」によって説明している<sup>29)</sup>。この立場では、未知の景観を既知の景観と比較照合を行うときに物体を回転させる (心的回転: mental rotation<sup>27)</sup>) ような変換を実行して最も近い既知の景観に正規化 (normalize) することによって認識が行われると考えられる。

一方、三次元構造ベースの物体表現理論によれば、適切に構造表現が抽出できれば回転コストはかからないことが予想される。Biederman と Gerhardstein は単一ジオンの比較照合課題を用いて、角度差をもって呈示される単一ジオンの認識において回転コストがかからないことを示した<sup>4)</sup>。また、彼らは以下のような条件が満たされた場合に視点不変な物体認知が可能となるとしている。

- ① 物体がジオンのようなパーツに分解可能であること
- ② 異なる物体はそれぞれ異なる構造表現 (GSD) で表現されていること
- ③ 異なる視点にわたって同一の構造表現が活性化されること

例えば、Bülthoff と Edelman の用いた刺激はこれらの条件が満たされていないため、視点依存的なパフォーマンスを示してしまうと考えられるが、これらの条件を満たす物体であれば例え新奇な物体であっても視点不変な物体認知が達成される。

しかし、Hayward や Tarr らは、Biederman と Gerhardstein が用いた刺激と同じ刺激を用いた追試を行い、単一ジオンの認識においても回転コストがかかることを示している<sup>19),30)</sup>。ただ、両グループが報告している実験協力者の反応時間や誤警報率に開きがあることから、実験協力者の課題に対する構えの違いが結果に影響を与えている可能性もある。

また、それぞれの理論は単に物体の視覚的特徴における異なる特性に焦点を置いている可能性も考えられる。Biederman と Bar は二つの物体を経時的に呈示し、それらの比較照合課題を実験協力者に行わせた<sup>5)</sup>。その際、二つの物体の視覚的特徴の量的 (metric) な特性 (例

えば複数のパーツが結合する角度)が異なる条件と、質的 (qualitative) な特性 (パーツの種類) が異なる条件間で比較を行った。その結果、質的な特性の変化に対する感度が量的な特性に対する感度より高く、視点の変化による影響も小さいことから、質的な特性、すなわち、構造表現に基づく物体表現の優位性を主張した。一方で、Foster と Gilson は、新奇物体の質的な特性と量的な特性を変化させた刺激を用いて心理実験を行った<sup>12)</sup>。その結果、質的な特性による認識コストと、量的な特性による認識コストが互いに独立で加算的な関係によって実験結果を説明できることを示した。よって、物体の質的な特性と量的な特性は分離して表象されている可能性がある。今後の物体認知の心理学研究ではこれらの特性について十分統制した実験を行って検討していく必要があるだろう。それにより、二つの理論を統合したモデルの構築も期待できる。

## ■S3 群 - 2 編 - 8 章

### 8-3 三次元物体認知にかかわる神経基盤

(執筆著: 笹岡貴史) [2008年5月 受領]

#### 8-3-1 生理学的知見

従来、物体認知には初期視覚野から高次視覚野を経て側頭葉に至る、腹側経路 (ventral pathway) と呼ばれるルートがかかわっていることが示唆されている<sup>32)</sup>。例えば、側頭 - 後頭葉に障害を受けると日常的な物体や顔などの認識ができなくなることが報告されている<sup>10)</sup>。また、サル IT 野 (inferior temporal cortex) の電気生理学的な測定により、IT 野の細胞はある程度複雑な視覚的特徴に対して選択的に反応し、類似した視覚的特徴に対して選択性をもつ細胞がコラム状の構造をなしていることが示されている<sup>14),28)</sup>。

Logothetis ら<sup>22)</sup>は、Bülthoff と Edelman が用いたペーパーリップオブジェクトを認識する課題をサルに学習させた後、同じ課題を遂行しているサルの IT 野のニューロンの活動を測定した。その結果、ペーパーリップオブジェクトのある景観に対して選択的に活動するニューロンが見つかった。そのようなニューロンが最もよく活動する景観から約±30° の範囲内の景観に対して活動することから、このようなニューロンの特性は、画像ベースの脳内表現を強く支持する証拠とされている。また、Riesenhuber と Poggio は、この IT ニューロンの特性をシミュレート可能な数理ネットワークモデルを提案している<sup>24)</sup>。

Booth と Rolls は単純な三次元物体を刺激として用いた実験により、サルの IT 野に視点依存的に反応する細胞と視点不変な反応をする細胞の両方を発見している<sup>9)</sup>。彼らは、サルの飼育ケージに新奇な物体を 10 個入れ、数週間実際にそれらに触れて遊ばせた後、これらの物体を含む様々な物体の画像をサルに呈示したときの IT ニューロンの活動を測定した。その結果、測定した約半数のニューロンは特定の view または特定の視覚的特徴に選択的な反応を示したが、14%のニューロンは同一の物体の異なる view に対して同じような反応を示す視点非依存的なニューロンであった。この視点非依存的なニューロンは視覚的特徴が類似した別の物体に対しては反応が小さいため、単純な特徴検出器でないことが示唆される。また、情報理論に基づく分析によって、これらのニューロンの発火パターンはそれぞれが一つの物体を表現する情報量をもっているというより、これらのニューロン群の発火パターンの組み合わせで、異なる物体が表現されていることが示された。以上の結果から、IT 野における視点非依存的ニューロンの反応は、ほかの視点依存的なニューロン群の反応の組み合わせによって形成されており、脳にそれぞれの物体に対応する個別のニューロンが存在するという、いわゆる「おばあさん細胞」のようなローカルな表現ではなく、分散表現になっていると考えられる。

#### 8-3-2 三次元物体認知の脳機能イメージング

1990 年代以降、脳機能イメージング技術の発達により、健康者の脳機能測定による三次元物体認知に関するデータが蓄積されてきている。本節 8-3-1 項で述べたような生理学的研究で示されてきたように、ヒトの fMRI (functional magnetic resonance imaging) 研究によっても物体を観察したときに活動する部位が発見されている。Grill-Spector ら<sup>17)</sup>は、fMRI adaptation と呼ばれる手法によって、サルの IT 野に対応する領域とされている LOC (lateral occipital complex) と呼ばれる領域の活動を調べた。fMRI adaptation とは、ある視覚刺激を提示したと

きに活動する脳部位が、同じ刺激を繰り返し提示することによって「順応」(adaptation)を起こし、fMRI 信号が減少するという現象のことをいう。例えば、ある刺激に対して、このような現象の起こった部位が存在すれば、まさにその刺激に対して選択的に活動する部位と考えられる。Grill-Spector らは、自動車や顔といった各カテゴリの画像を繰り返し提示することで、それらに対して選択的に活動が減少する領域が存在することを示した。また、LOC のなかでも特に腹側前部の領域 (LOa/PF) は、同じ物体の画像平面上での位置を変えた刺激を呈示したり、サイズを変えた刺激を呈示したりすることでも、変化のない場合と同様に活動の現象が起こることから、LOC には位置や大きさに対して不変な反応をする領域があることが示唆された。

Vulliamier ら<sup>34)</sup>は、様々な物体の異なる view を実験協力者に呈示し、反復ブライミング課題を行っている実験協力者の脳活動を fMRI で測定した。その結果、左半球の下側頭葉にある紡錘状回 (fusiform gyrus) という領域で、view にかかわらず同じ物体が呈示されたときに活動が減少する傾向がみられたのに対し、右の紡錘状回では同じ view が繰り返し呈示されることで活動が減少する傾向が見られた。以上のことから、下側頭葉では左右半球でそれぞれ視点非依存性、視点依存性の物体表現が行われていることが示唆される。

比較照合された物体は適切な物体クラスに分類される必要があるが、このような機能は前頭葉が担っていると考えられる。例えば、Bar らは物体を観察したときに後頭葉・側頭葉にその物体が何であるかという候補をあげるトップダウンの処理にかかわっており、瞬時の物体認識を可能にしていることを示唆している<sup>1)</sup>。また、Freedman らはサルの外側前頭前皮質に特定のカテゴリの物体に選択的に反応するニューロンを発見している<sup>13)</sup>。このように、物体を知覚し、物体の脳内表現との比較照合を行い、何らかのカテゴリへとカテゴリ化するという物体認知過程は、後頭葉から側頭葉、そして前頭葉を含む広範囲のネットワークに基づいていると考えられる。

## ■S3 群 - 2 編 - 8 章

### 8-4 今後の展望

(執筆著：笹岡貴史) [2008年5月 受領]

本章 8-3 節であげてきたように、これまでの物体認識にかする脳研究では、物体の内部表現が脳のどこに存在し、それがどのような特性をもっているのかということについて主に研究がなされてきた。しかし、実際に物体認知を行うときにそれがどのように入力画像と比較照合が行われているかについて調べた研究は存在するが（例えば 16), 33)) いまだに少なく、比較照合過程の脳内メカニズムについてはまだ明らかになっていない。

また、近年、物体の認知は視覚系だけではなく、運動系とも密接に結びついていることを示唆する研究が行われている。Harman ら<sup>18)</sup>は、新奇な三次元物体を実験協力者に学習させる際、一方のセットの物体についてはトラックボールを手で操作することで能動的に物体の view を観察して学習させ、他方のセットの物体についてはほかの実験協力者が能動的に物体を操作したリプレイを観察して学習させた。その後に行った再認課題で、能動的に学習した物体セットに対する再認反応時間が、受動的に学習した物体セットよりも有意に短くなった。この結果は、運動系からの情報が未知の景観を既知の景観に正規化する際の物体イメージの変換、もしくは、物体の内部表現の獲得の促進に寄与している可能性を示唆している。

今後、心理学的研究、脳研究の相補的な進展によって、比較照合過程、物体の内部表現の獲得過程を含む新たな三次元物体認知のモデルが提案されることが期待される。

#### ■参考文献■

- 1) M. Bar, K.S. Kassam, A.S. Ghuman, J. Boshyan, A.M. Schmid, A.M. Dale, M.S. Hämäläinen, K. Marinkovic, D.L. Schacter, B.R. Rosen, and E. Halgren, "Top-down facilitation of visual recognition," *Proc. Natl. Acad. Sci. USA*, vol.103, pp.449-454, 2006.
- 2) I. Biederman and E.E. Cooper, "Size invariance in visual object priming," *J. Exp. Psychol. Hum. Percept. Perform.*, vol.18, pp.121-133, 1992.
- 3) I. Biederman, "Recognition-by-components: A theory of human image understanding," *Psych. Rev.*, vol.94, pp.115-147, 1987.
- 4) I. Biederman and P.C. Gerhardstein, "Recognizing depth-rotated objects: Evidence and conditions for 3D viewpoint invariance," *J. Exp. Psychol. Hum. Percept. Perform.*, vol.19, pp.1162-1182, 1993.
- 5) I. Biederman and M. Bar, "One-shot viewpoint invariance in matching novel objects," *Vision Res.*, vol.39, pp.2885-2899, 1999.
- 6) M.C. A.Booth and E.T. Rolls, "View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex," *Cereb Cortex*, vol.8, pp.510-523, 1998.
- 7) H.H. Bülthoff and S. Edelman, "Psychophysical support for a two-dimensional view interpolation theory of object recognition," *Proc. Natl. Acad. Sci. USA*, vol.89, pp.60-64, 1992.
- 8) M. Dill and M. Fahle, "Limited translation invariance of human visual pattern recognition," *Percept Psychophysics*, vol.60, pp.65-81, 1998.
- 9) S. Edelman, "Computational theories of object recognition," *Trends Cogn. Sci.*, vol.1, pp.296-304, 1997.
- 10) M.J. Farah, "Visual Agnosia: Disorders of Object Recognition and What They Tell us about Normal Vision," MIT Press, Cambridge, 1990.
- 11) J. Fiser and I. Biederman, "Invariance of long-term visual priming to scale, reflection, translation, and hemisphere," *Vision Res.*, vol.41, pp.221-234, 2001.
- 12) D.H. Foster and S.J. Gilson, "Recognizing novel three-dimensional objects by summing signals from parts and

- views,” *Proc. R. Soc. Lond. B*, vol.269, pp.1939-1947, 2002.
- 13) D.J. Freedman, M. Riesenhuber, T. Poggio, and E.K. Miller, “Categorical Representation of Visual Stimuli in the Primate Prefrontal Cortex,” *Science*, vol.291, pp.312-316, 2001.
  - 14) I. Fujita, K. Tanaka, M. Ito, and K. Cheng, “Columns for visual features of objects in monkey inferotemporal cortex,” *Nature*, vol.360, pp.343-346, 1992.
  - 15) C. Furlanski and S.A. Engel, “Perceptual learning in object recognition: object specificity and size invariance,” *Vision Res.*, vol.40, pp.473-484, 2000.
  - 16) I. Gauthier, W.G. Hayward, and M.J. Tarr, “BOLD activity during mental rotation and viewpoint-dependent object recognition,” *Neuron*, vol.34, pp.161-171, 2002.
  - 17) K. Grill-Spector, T. Kushnir, S.E. Delman, G. Avidan-Carmel, Y. Itzhak, and R. Malach., “Differential processing of objects under various viewing conditions in the human lateral occipital complex,” *Neuron*, vol.24, pp.187-203, 1999.
  - 18) K.L. Harman, G.H. Humphrey, and M.A. Goodale, “Active manual control of object views facilitates visual recognition,” *Curr. Biol.*, vol.9, pp.1315-1318, 1999.
  - 19) W.G. Hayward and M.J. Tarr, “Testing conditions for viewpoint invariance in object recognition,” *J. Exp. Psychol. Hum. Percept. Perform.*, vol.23, pp.1511-1521, 1997.
  - 20) J.E. Hummel and I. Biederman, “Dynamic binding in a neural network for shape recognition,” *Psych. Rev.*, vol.99, pp.480-517, 1992.
  - 21) J.E. Hummel, “Where view-based theories of human object recognition break down: the role of structure in human shape perception,” in *Cognitive Dynamics: conceptual change in humans and machines*, ed. E. Dietrich and A. Markman, chapter 7, Erlbaum, Hillsdale, NJ, 2000.
  - 22) N.K. Logothetis, J. Pauls, and T. Poggio, “Shape representation in the inferior temporal cortex of monkeys,” *Curr. Biol.*, vol.5, pp.552-563, 1995.
  - 23) D. Marr and H.K. Nishihara, “Representation and recognition of the spatial organization of three-dimensional shapes,” *Proc. R. Soc. Lond. B*, vol.200, pp.269-294, 1978.
  - 24) M. Riesenhuber and T. Poggio, “Hierarchical models of object recognition in cortex,” *Nat Neurosci*, vol.2, pp.1019-1025, 1999.
  - 25) I. Rock and J. DiVita, “A case of viewer-centered object perception,” *Cognit Psychol.*, vol.19, pp.280-293, 1987.
  - 26) T. Poggio and S. Edelman, “A network that learns to recognize three-dimensional objects,” *Nature*, vol.34, pp.263-266, 1990.
  - 27) R.N. Shepherd and J. Metzler, “Mental rotation of three-dimensional objects,” *Science*, vol.171, pp.701-703, 1971.
  - 28) K. Tanaka, H.A. Saito, Y. Fukada, and M. Moriya, “Coding visual images of objects in the inferotemporal cortex of the macaque monkey,” *J Neurophys.*, vol.66, pp.170-189, 1991.
  - 29) M.J. Tarr, “Rotating objects to recognize them: A case study of the role of viewpoint dependency in the recognition of three-dimensional objects,” *Psychon Bull Rev*, vol.2, pp.55-82, 1995.
  - 30) M.J. Tarr, P. Williams, W.G. Hayward, and I. Gauthier, “Three-dimensional object recognition is viewpoint dependent,” *Nat Neurosci*, vol.1, pp.275-277, 1998.
  - 31) S. Ullman, “Aligning pictorial descriptions: an approach to object recognition,” *Cognition*, vol.32, pp.193-254, 1989.
  - 32) L.G. Ungerleider and M. Mishkin, “Two cortical visual systems,” in *Analysis of visual behavior*, ed. D.J. Ingle, M.A. Goodale, and R.J. W. Mansfield, MIT Press, Cambridge, 1982.
  - 33) J. Vanrie, E. Beatse, J. Wagemans, S. Sanaert and P. Van Hecke, “Mental rotation versus invariant features in object perception from different viewpoints: an fMRI study”, *Neuropsychologia*, vol.40, pp.917-930, 2002.
  - 34) P. Vuilleumier, R.N. Henson, J. Driver, and R.J. Dolan, “Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming,” *Nat Neurosci*, vol.5, pp.491-499, 2002.