

スケールと回転に対する不変な特徴対応付けに基づく神経メカニズム

A Correspondence-based Neural Mechanism for Scale and Rotation Invariant Feature Processing

佐藤能臣 (PY), Jenia Jitsev, Christoph von der Malsburg

Yasuomi D. Sato(PY), Jenia Jitsev and Christoph von der Malsburg

Frankfurt Institute for Advanced Studies (FIAS), Johann Wolfgang Goethe University

sato@fias.uni-frankfurt.de

Abstract — We present a neurally plausible mechanism for finding correspondences between two images of the same object, taken under a natural condition at different unknown relative scale and/or rotation. The mechanism is based on a macrocolumnar cortical model and is able to detect the most likely relative scale and rotation by comparing features extracted at points near the center of each image, by competitive evaluation of transformation-specific similarities.

Keywords — Invariant Object Recognition, Feature Correspondence, Global Signature, Transformation Specific Similarities, Macrocolumnar Model

1 導入

物体が回転したり、大きさを変えたりすると、それに従い、目に映る物体の像も変わっていくが、我々人間の脳は、その大きさや回転の角度に依らず、瞬時にその物体であると認識する。このような認識過程において、大きさや回転の変換状態に関する視覚情報佚われてしまい、目に映る物体の像の表現を著しく曖昧にさせてしまっている。このような曖昧性があるにも関わらず、どの大きさからもどの回転からも正しく認識する仕組みを解明していくことは、我々人間の脳と意識の関係を本質的に理解する上で、非常に重要な問題であると言えよう。

このような人間の脳が持つ賢さからの脳型情報処理の仕組みを解明していくための、物体同定認識システムのモデル化では、脳内に個別に物体の像の表現が既に記憶として貯蔵されているものとし、それらの貯蔵されている像の表現と目に映る像の表現とが正しくマッチングをすることで初めて、その物体であると認識するとした [1]。実際、記憶されている像の表現に対して、目に映る像の表現にはどのような回転/大きさの変換が加わっているのかを予め知ることは出来ない。先行研究における同定認識システムのモデル化では、このような自然な不確定条件は捨て去られ、観測者自身が自分の目に映る像の回転変換やサイズ変換を既知しているという大きなジレンマを抱えた仮定の下で、認識システムのモデル化が行なわれてきた [2]。

本研究では、Global Signature と呼ばれる手法を用い、Gabor ウェーブレット変換によって画像の中心

点から抽出される特徴量の空間周波数要素/方位要素の数によって定められる、スケール/回転変換に関する離散パラメータを認識システムに内在的に定義していくことで、入力画像に加わる変換を推定し、従来の不自然な仮定を破棄し、目に入る時点で物体の像の回転・スケール変換状態が不確定であるというより人間らしい条件のもとでも、その変換状態を検知しつつ、同時に、物体同定を可能にする認識システムのモデル化を行った。

2 類似性を利用した変換状態の特定

本研究における認識システムでは、まず初めに、入力画像 (I) と貯蔵画像 (M) を用意する。画像 I と画像 M 上の物体は同一のものであるが、画像 M 上の物体には、画像 M に対して任意の回転・スケール変換が施されているものとする。Gabor ウェーブレット変換によって、画像 I と画像 M それぞれの中心点から、 $(m_a + m_r + m_\theta)$ 個と m_r 個の空間周波数要素と n 個の方位要素で構成される特徴量を抽出する [3][4]。ここで、 m_r を空間周波数に関する基本要素数とすると、 m_a と m_θ は、それぞれ実際の画像 M の物体に対する画像 I の物体のスケール・ダウンとスケール・アップのステップ数に相当し、その物体のサイズは、Gabor フィルター内の空間周波数係数に依る。

次に、Global Signature により、画像 I と M の特徴における、個々の空間周波数要素について、もう片方の要素である方位要素 n 個の平均をとり、空間周波数のグループ・ベクトルの 1 つの要素を作っていく。このような平均は、個々の方位要素に対してもなされ、結果として、両画像の中心点から抽出された特徴量は、空間周波数特性と方位特性の 2 つのグループ・ベクトルに分解されることになる。

空間周波数特性に関するグループ・ベクトル I と M の要素間の比較の仕方により、 $(m_a + 1 + m_\theta)$ 個のスケール変換が離散パラメータ s として与えられ、方位特性のグループ・ベクトルについても同様に、 n 個の回転変換がパラメータ r としてシステム内に定義され、これらのスケール (S) / 回転 (R) 変換パラメータ s や r それぞれについて、グループ・ベクトル I と M 間の類似性が計算される：

$$S_s^s(\bar{J}^I, \bar{J}^M) = \frac{\sum_l J_{l+r}^I J_l^M}{\sqrt{\sum_l (J_l^I)^2 \sum_l (J_l^M)^2}},$$

$$S_r^R(\bar{J}^I, \bar{J}^M) = \frac{\sum_k J_{k-r+n}^I J_k^M}{\sqrt{\sum_k (J_k^I)^2 \sum_k (J_k^M)^2}}, \quad (k-r) \bmod n$$

ここで、 J は、画像とMの中心点から抽出された特徴量の要素を表す。このように、内在的に与えられる全ての変換パラメータに対する類似性が計算され、画像Iの変換状態が予め分からなくても、類似性が最大値をとるような変換パラメータが、最も見込まれる画像Iの変換状態に相当することになる。

3 皮質コラム・ダイナミクスによる変換検知

前章では、画像の中心点から抽出された特徴の性質をうまく利用し、内在的に離散変換パラメータを定義した。次に、スケール変換 (S) 或いは、回転変換 (R) それぞれに対して、その変換パラメータ s 、或いは r を表すミニコラムの電位度 (\bar{v}) 集合 (皮質コラム) の非線形ダイナミクス[5] :

$$\frac{d\bar{v}_\gamma^C}{dt} = f_\gamma(\bar{v}) + \kappa S_\gamma^C(\bar{J}^I, \bar{J}^M), \quad (C, \gamma) \in ((S, R), (s, r))$$

を適用することで、離散パラメータ以外のスケール/回転変換状態に対しても柔軟な認識を実現した。

この皮質コラム・ダイナミクスでは、ミニコラム間の相互作用項を含む関数 f と類似性 s の入力値により、類似性が最大値をとるような変換パラメータのミニコラムは、その活動を維持し、それ以外の変換パラメータに関するミニコラムについては、その活動を停止することで、画像Iの物体の関連するスケール変換状態と回転変換状態を検知する。

この皮質コラムモデルでは、内在的に与えられた変換パラメータを中心に、ある特定の変換領域では、その与えられた変換パラメータの類似性は常に最大値をとり続けるため、その変換パラメータのミニコラムの活動は、その変換領域では活動停止が起らない。その関連する変換領域をミニコラムがカバーするように認識システムに記憶させることで、入力画像の物体にどのような変換状態が加わっているのかを予め知らなくても、もっとも見込まれるスケール/回転変換を検知しつつ、その物体をその物体として同定するシステムを実現した。

4 議論

本研究では、Gaborフィルターの空間周波数特性の要素と方位特性の要素から、見込まれる画像のスケール変換や回転変換を離散的にパラメータ化し、その変換パラメータの検知機能を認識システムに取り入れることで、貯蔵されている物体の表現に対して、どのようなスケール変換、或いは、どのような回転変換が入力画像に組み込まれているのかを予め知ることなく、

広い変換領域で、関連する変換情報を検知するシステムを開発した。

このような内在的パラメータ化のために必要であったGlobal Signatureによる特性分解は、Subsystem Coordinationと呼ばれる、これまでの階層型の視覚情報処理システムとは異なる、興味深い視覚情報処理の仕組みを提案する[5]。

これは、皮質コラム・ダイナミクスが、物体の変換状態に関する視覚情報、物体の同定に関する視覚情報、物体の分類に関する視覚情報をそれぞれ断片的に、かつ独立に保持していると考え、そのような複数の様々なコラム活動から発せられる断片的な視覚情報が相互に影響を及ぼし合い協調していくことで、初めて1つの大きな視覚情報が構築され、その物体であると認識していく仕組みである。

この仕組みを利用したシステム構築では、膨大な数の情報次元を同時処理する必要があり、議論する余地が大いにあるものの、更により汎用性の広い視覚認識システム構築も見込まれる。例えば、これまでの不変同定認識システムのモデル化では、物体の大きさや回転だけではなく、観察者が見る位置をわずかに変えることによって生じる、目に映る物体の像の位置ずれに対して不変な認識モデルについても盛んに議論されてきた。

このような知見も踏まえ、今後は、入力画像の大きさ、回転、位置ずれに関する視覚情報を検知していくと同時に、物体を同定したり、識別したりするシステムのモデル化を行っていく。このような研究は、実際の生活場面における我々人間の視覚認識を説明していく上で非常に意義深いものであると言えよう。

謝辞

本研究は、Hertie FoundationとEU project “Daisy” FP6-2005-015803の支援を受けました。本研究課題を達成するにあたり、多大な御助言・御指導を頂いたFIASニューログループのメンバーに深く感謝します。

参考文献

- [1] N.K. Logothetis, J. Pauls, T. Poggio (1995) *Current Biology*, **5**(5), 552-563.
- [2] M. Lades (1995) Ph.D Thesis of Ruhr-Univ. Bochum.
- [3] D. Gabor (1946) *J.IEE*, **93**, 429-459.
- [4] L. Wiskott and C. von der Malsburg (1996) “Lateral Interactions in the Cortex: Structure and Function” (eds. J. Sirosh, R. Miikkulainen, Y. Choe), Chapter 11.
- [5] J.Lücke and C. von der Malsburg (2004) *Neural Computation*, **16**, 501-533.
- [6] Y.D. Sato, J. Jitsev, C.von der Malsburg (2008) *Proc. ICANN2008*, (accepted).