

軌道アトラクタを用いた視点に依存しない物体認識のモデル

A model for view-invariant object recognition using trajectory attractors

小林 遼太 (PY)[†], 山根 健[†], 森田 昌彦[†]

Ryota Kobayashi(PY), Ken Yamane, and Masahiko Morita

[†] 筑波大学大学院システム情報工学研究科

ryo@bcl.esys.tsukuba.ac.jp

Abstract— We propose a biologically-inspired model for object recognition based on distributed representations and continuous state transitions of a recurrent neural network. Using this model, we built a system that can recognize images of 30 objects taken from various angles without matching with templates.

Keywords— **View-invariant Recognition, Non-monotone Neural Network**

1 はじめに

一般に、物体の見え方は環境や見る角度によって大きく変化するため、現実世界において多数の物体を識別することは、機械にとって非常に難しい。単純なプレートマッチングでは、一つの物体の無数の見えに対応したプロトタイプが必要となるため、物体が増えるにつれて記憶すべきデータ量も認識に要する時間も飛躍的に増大してしまうからである。一方、人間や高等動物は、知っている物体であれば、ほとんどどんな視点からでも正しく認識することができる上に、記憶している物体の数がいくら増えても認識に要する時間はほとんど変わらない。

近年の研究では、人の脳における物体の表現は、3次元の形状を表すというよりも、見え方に依存する Appearance-based なものであるという考えが有力であり、これに基づく様々なモデルが提案されている [1]。しかし、もしそのような表現に基づき、かつ各見えに対応するプロトタイプを記憶しているのでないとすると、かなり複雑な境界面を構成しなければならない。例えば A と B の 2 人の顔を識別する場合、A の横顔は、明らかに A の正面の顔よりも B の横顔に似ているからである。

本研究では、高次元の特徴空間を非単調神経回路網 [2] の状態空間とし、そこに線状のアトラクタを形成することによって、多数のプロトタイプを直接記憶することなくパターン識別を行うシステムを構築し、基礎的実験を行った。

2 軌道アトラクタを用いた物体認識モデル

2.1 特徴の抽出と符号化

前処理として画像から特徴を抽出する際、従来の方法では、一般にできるだけ少数の特徴的な（特定の物体だけに含まれるような）特徴量を選ぶことが求められた。これに対し、本モデルでは、多数の特徴量に対象物体の情報が広く含まれ、見えの変化に対して特徴ベクトルが徐々に変化するようにする必要がある。今回はとりあえず以下のような処理を行ったが、これが最適というわけでは決していない。

まず、抽出する特徴は、大きく分けて形状に関する特徴（エッジ情報）と色に関する特徴（色ヒストグラム）の 2 種類とした。エッジ情報は、カラー画像を白黒画像に変換した上で方位選択性を持つ Gabor Filter をかけた後、位置ずれをある程度許容するためにぼかし操作を行って取得する。なお、周辺部は中心部より 3 倍ほど大きくぼかすこととした。色ヒストグラムは、入力画像の中心部領域を HSV 系に変換し、そこから明度にあたる V を除いた 2 次元の値の分布を計算することによって得る。

各特徴量（エッジ 493 個、色 100 個）は 3 段階に量子化した上で、各段階を $[-1, -1, 1, 1]$, $[-1, 1, 1, -1]$, $[1, 1, -1, -1]$ で表すことによって符号化する。これは、特徴量が段階的に変化したとき、符号パターンも段階的に変化するようにするためである。以上により得られた 2192 次元の 2 値パターンを、元の画像の特徴ベクトルとする。

2.2 学習と認識

ある物体を徐々に回転したとき、上記の特徴ベクトルは 2192 次元のパターン空間において連続的な軌道を描くが、このような軌道は非単調神経回路網に軌道アトラクタとして埋め込むことができる [2]。そこで、ある物体をさまざまな角度から見た画像がある場合に、それぞれの特徴ベクトルを補間した軌道をいくつか作成し、2192 個の素子からなる非単調神経回路網に学習させる。このとき、軌道の終点はその物体の代表的な画像（正面画像）の特徴ベクトルとする。これをすべての物体について何度か繰り返すと、同じ物体の特徴ベクトルを結んだ軌道アトラクタ群が形成される（図 1）。

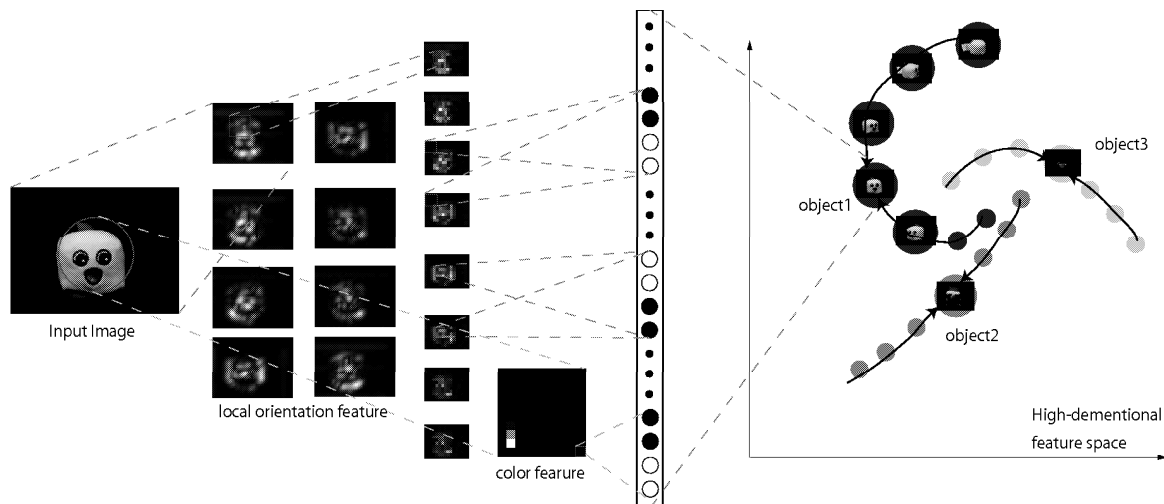


図 1: 特徴抽出過程とモデルの状態空間

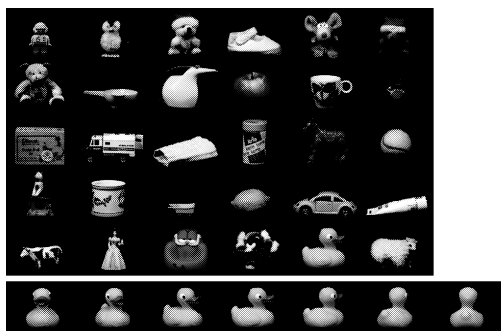


図 2: 学習に用いた画像群

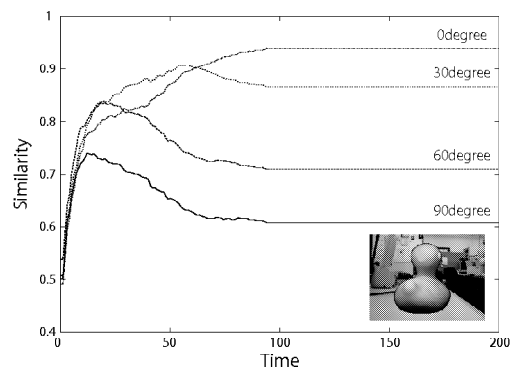


図 3: 認識の過程

学習後の回路網にある軌道アトラクタの近くに位置する特徴ベクトルを初期状態として与えると、回路網の状態はその軌道に引き込まれた後、軌道に沿って遷移し、正面画像の特徴ベクトルに達することが期待される。これにより、最初に与えた画像がどの物体の画像であるか識別することができる。

3 実験

実験には、図 2 に示す 30 個の物体に関する画像群を用いた。これらは、Amsterdam Library of Object Images の Image-Database より取得したものである。それぞれの物体について、観察方向が水平方向 -90 度から 90 度まで 30 度ごとの画像 7 枚を用いて軌道を生成し、学習を行った。

学習後のシステムに、学習には用いていない、 -80 度および 40 度方向の画像 60 枚を入力したところ、すべて正しく識別することができた。また、 70 度方向の画像に背景を合成した画像を入力した場合も、正しく認識することができた。認識の過程を示したのが図 3 であり、回路網の状態ベクトルと、学習に用いた画像の特徴ベクトルとの類似度を時間変化を示している。このように、入力画像と学習に用いた画像の特徴ベクトルの類似度

が最大 0.5 程度であっても、正面画像の特徴ベクトルがほぼ正確に想起されていることがわかる。

4 おわりに

軌道アトラクタを用いて、大きく見え方の異なる同一物体の画像を同一のパターンとして認識するモデルを提案した。このモデルは、構造も学習則も単純であり、複雑な統計的計算を必要としない。また、認識に要する時間が観察角度に多少依存する一方で、記憶する物体の数にはほとんど依存しないなど、従来のモデルより人間に似ていると思われる。

今後は、より多くの物体を用いた実験を進めると共に、特徴抽出の方法や色情報の扱い方の改善、文脈修飾の導入などにより、認識能力の向上を図りたい。

参考文献

- [1] Roth Peter M., Winter Martin. (2008) "Survey of appearance-based methods for object recognition." Technical Report, **ICG-TR-01/08**.
- [2] 森田昌彦 (1997) "非単調神経回路網による時系列パターンの認識." 信学論, **D-II-J81**, 7, 1679-1688.