

ニューラルネットによる輪郭抽出

- 多層マルコフ確率場モデル -

Contour extraction by neural networks
- Multi-layered Markov random field model -

本郷節之 川人光男 乾 敏郎
Sadayuki Hongo Mitsuo Kawato Toshio Inui

ATR視聴覚機構研究所
ATR Auditory and Visual Perception Research Laboratories

1. はじめに

視覚情報処理の役割は、網膜上に投影された2次元画像から外界の3次元情報を推定することにある。この問題は一般に不良設定問題(ill-posed problem)であり、唯一の解を求めることはできない。けれども我々は、通常唯一の視覚像を持っている。したがって、これらの不良設定問題を解く際に、何らかの手段を用いて欠落している情報を推定していることが考えられる。

Harr¹⁾は、この推定を行う過程で、拘束条件として外界の物理的な法則を利用していると考えた。すなわち、我々は、外界の状態が明るさや色、奥行きなどへ及ぼす影響に関する先見的知識を持っており、網膜像から外界の状態を推定する際に、観測された網膜像とこの知識とから最も確からしい外界の状態を推測するということである。

この思想は、Poggioら²⁾によって、評価関数を最小化する解を求める問題、「標準正則化理論(standard regularization theory)」としてまとめ上げられた。ここでは3次元情報の復元のみならず、輝度エッジの検出のような明るさ情報の処理も、“まばら”な入力情報から輝度分布を推定する不良設定問題として捉えている。

さらにGeman and Geman³⁾によって導入された、強度過程(intensity process)の不連続を表す線過程(line process)は、不連続を有する輝度分布の推定を定式化することを可能にした。彼らは、ノイズに汚されたデータから原画像を推定し復元する、画像復元の問題を取り扱ったが、その基本的なアイデアは初期視覚のモデルにも適用できる。彼らは、画像復元の問題を、観測画像から最も確からしい過去を推定する問題、すなわち最大事後確率(MAP)推定と捉えた。そして画像のマルコフ確率場(MRF)

モデルが、Gibbs分布のエネルギーを定義することと等価であることを示し、MAP推定がMRFのエネルギー最小化によって達成できることを示した。さらにこのエネルギー最小化は、模擬された焼きなまし(simulated annealing)を伴う統計的緩和法(stochastic relaxation)によって厳密に達成できることを示した。

筆者ら^{4,5)}は、この統計的緩和法を、自然画像の輪郭線を推定する問題に適用した。ここでは、輪郭線の構造に関する先見的知識であるエネルギーパラメータを教師輪郭線から獲得する、学習アルゴリズム⁶⁾が用いられている。

本稿では、まずMRFモデルおよびエネルギー学習アルゴリズムを紹介し、般化能力に関する検討、および大局的構造表現の利用に関する検討について述べる。

2. MRFモデル

Geman & Geman によって提案された結合型MRFモデルは、各画素の間に線過程と呼ばれる画像の不連続を表現する仮定の確率過程を導入し、その相互作用を用いて画像復元や輪郭線抽出を行うことができる。モデルは、近傍系における線過程の連結の形式を、その出現確率を反映する局所的エネルギーとして持つことができる。例えば、輪郭線の途切れを禁止するといった局所的な知識を持たせて、連結性に優れた輪郭線抽出を行わせることができる。

強度過程を入力画像の濃淡値に固定したまま、線過程のみを変化させて輪郭線を抽出することを考える。いま、図1の様な近傍系を考えてみよう。1, -, 0 はそれぞれ縦、横の線過程と強度過程を表す。×印をつけたサイトを中心に考える。このサイトが取り得る状態は、そこに線過程が存在する($l_{ij} = 1$)、または存在しない($l_{ij} = 0$)のいずれかである。そこで、

画像の濃度勾配のポテンシャルを $U_0(\omega)$ ，線過程のポテンシャルを $U_1(\omega)$ ，サイト全体のポテンシャルを $U(\omega)$ とすると次式のようになる。

$$U(\omega) = U_0(\omega) + U_1(\omega)$$

$$U_0(\omega) = V_0 \cdot \sum_{i,j} \{ (g_{i,j+1} - g_{i,j})^2 \cdot (1 - h_{i,j}) + (g_{i+1,j} - g_{i,j})^2 \cdot (1 - v_{i,j}) \}$$

$$U_1(\omega) = \sum_{C \in \mathcal{C}} V_C(\omega)$$

ただし、 $h_{i,j}$ および $v_{i,j}$ は各々横および縦の線過程の状態を表している。このポテンシャルに従って十分に状態変化を繰り返すことによって、画像の境界として輪郭線を求めることができる。

3. 学習アルゴリズム

画像の MRF モデルを特徴付ける局所的エネルギーを、真の画像の状態を教師信号として同じ統計的性質を持つ画像データに対して推定・学習するアルゴリズムが川入ら⁶⁾によって提案されている。濃淡画像から輪郭線画像を推定するのに必要となる局所的エネルギー V_{ξ_i} の値を、目標とする輪郭線画像を教師信号として学習するアルゴリズムは、次式のように表すことができる。

$$V_{\xi_i} = -\eta \left\{ \sum_{C \in \mathcal{C}} I_i(\xi) - \sum_{C \in \mathcal{C}} I_i(\xi)' \right\} / \sum_{C \in \mathcal{C}} I_i(\xi)$$

ここで I_i はクリーク C の状態が ξ_i のときだけ 1 を取り、それ以外では 0 となる ξ_i の定義関数である。右辺第 1 項は、学習の目標となる輪郭線画像（教師画像）において、クリーク C が ξ_i をとる個数を画像全体で数え、それをサイトの数で平均したものである。一方右辺第 2 項は、ある局所的エネルギー V_{ξ_i} を固定した時の MRF モデルで、クリーク C が ξ_i をとる個数の平均値である。これは、MRF モデルで画像を生成して計算する。

図 2 に、学習前および 30 回学習後において抽出された輪郭線を示す。学習によってノイズが減少し、連結性が向上していることがわかる。

4. 般化能力の検討

学習アルゴリズムを議論するにあたって、般化能力は最も重要な項目のひとつである。例えば人間は通常、初めて目にするものであっても、その輪郭線を抽出することができる。すなわち、

本学習アルゴリズムにおいても、例からの学習によって獲得されたエネルギーパラメータが、学習した画像以外の（未知の）画像に対してどの程度の一般性を持つのかを調べる必要がある。

そこで、異なる 3 人の顔画像について、学習されたエネルギーパラメータの値を比較した。図 3 に学習の前後におけるパラメータの変化を示す。濃度勾配のエネルギーは、画像によって大きく異なっているので、 V_0 の値は濃度勾配の平均値の比によって求めている。また図中のパラメータの値は、各画像別に平均値からの比を用いて表してある。

図から、エネルギーパラメータの学習は、増加するものは増加、減少するものは減少と、全て同じ方向に進行していることがわかる。しかも、学習後の値は非常に類似している。このことから、本学習アルゴリズムには、ある程度の般化能力があると考えられる。

5. 多層 MRF モデル

輪郭線抽出の問題は、観測画像に基づいて外界の状態の不連続を推定することであり、初期視覚の重要な問題の一つである。MRF モデルでは局所的な相関のみを用いている。しかし人間は通常、複数の解像度を用いて外界を知覚していると考えられている^{7,8)}。したがって、局所的な情報と大域的な情報とを統合して外界の状態を知覚しているものと考えられる。

このような観点から、筆者らは、局所的な表現と大域的な表現を持ち、その間に相互作用を持つ多層 MRF モデルを提案している^{4,5)}。ここでは、異なる解像度を持つ複数の MRF として外界が表現されており、各層の間に局所的な結合を持っている。図 4 に、多層 MRF モデルの概念図を示す。

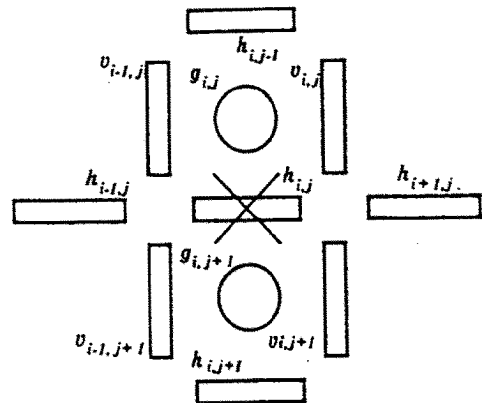


図 1 $h_{i,j}$ の相互作用近傍系

6. 多層MRFモデルによる輪郭抽出

図5は層間における線過程の配置の対応を表している。先に粗いレベル(レベル $n+1$)で輪郭線を求めておき、それよりも1段細かいレベル(レベル n)で輪郭線を求める際に、レベル $n+1$ でのクリークの要素と対応するクリーク要素のエネルギーが低くなるように条件づけて計算を行う。すなわち、0から $N-1$ までの N 個のレベルからなる階層的MRFモデルにおいて、レベル n におけるクリークを C^n ($n=0, 1, \dots, N-1$)、そのとり得る状態を ξ^n_i 、各 ξ^n_i に与えられる局所的エネルギーを $V^n(\xi^n_i)$ 、1または微小な値 ϵ をとる係数を A とすると、実際にエネルギー計算に用いる局所的エネルギー $V^n(\xi^n_i)$ は次式のように表すことができる。

$$V^n(\xi^n_i) = V^{n+1}(\xi^{n+1}_k) \cdot A$$

$$A = \epsilon \ll 1 \text{ if } \xi^n_i \in \Xi^n_j(\xi^{n+1}_k)$$

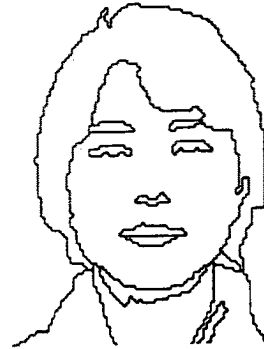
$$A = 1 \text{ otherwise}$$

$\Xi^n_j(\xi^{n+1}_k)$ はレベル $n+1$ における線過程の配置 ξ^{n+1}_k とうまく対応するレベル n での配置の集合である。この手法を用いれば、状態変化の局所性を保ちつつ、粗いレベルで求めた全体的構造が細かいレベルの輪郭線画像に反映されることになる。

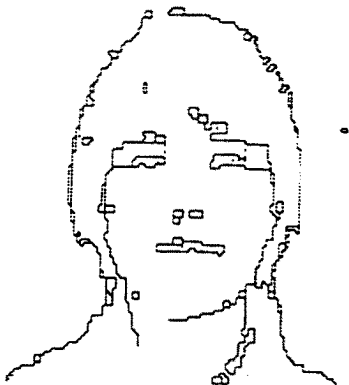
図6に単層および多層MRFモデルによって抽出した輪郭線画像を示す。(a)が単層モデルでの抽出結果、(b)が多層モデルの抽出結果である。なお、(c)は低解像度の単層モデルで抽出した輪郭線である。ここでは簡単のため、低解像度層の対応点に存在している局所構造(ending, crossing等)の出現確率を高く(



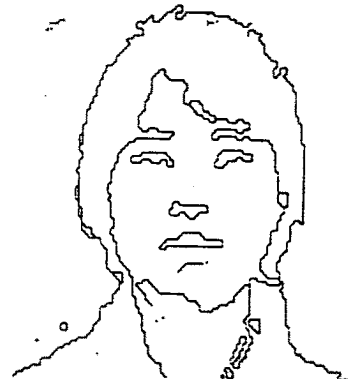
(a) 原画像



(b) 教師輪郭線



(c) 学習前



(d) 30回学習後

図2 学習による輪郭線抽出能力の変化

$V_c=0.75V_c$) している。さらに、前述の学習アルゴリズムを用いてエネルギーパラメータを学習させている。学習は30回繰り返した。

図から、単層モデルの出力には向かって右側の目の右上および画面の右上にノイズが見られるが、多層モデルの出力ではこれが消えており、さらに輪郭線の連結性も向上していることがわかる。していることがわかる。

7. まとめ

輪郭線抽出を目的とした結合型MRFモデル、学習アルゴリズム、および結合型MRFモデルの多層化について概説し、さらに般化能力および大局的構造の導入に関する検討について述べた。本学習アルゴリズムが般化能力を持つことを、実験的に確認することができた。さらに、2つの解像度からなる多層MRFモデルを用いて、大局的な形状を反映した輪郭線抽出を行うことができた。

8. おわりに

MRFモデルの多層化は、輪郭線の品質向上のみならず、計算速度の向上をも可能にすると考えられる。これは、MRFモデルに対して、より適切な条件を与えることによって、収束を早めることができるからである。

MRFモデルを多層化するにあたって、本稿では多重解像度表現を用いた検討を行っている

が、端点や交点などのような、さらに抽象化された構造的特徴との統合も重要な課題であると考えられる。乾ら^{9,10)}は、画像の特徴点を同時並列的に抽出する「特徴点検出アルゴリズム」を提案している。このアルゴリズムは、複数のサイズのエッジオペレータを組み合わせて、輪郭線において曲率の大きな部分、分岐点、端点、交点等を抽出することができる。このアルゴリズムを用いて特徴点を検出し、MRFモデルによって補間すれば、より高速に高品質の輪郭線抽出を行うことができると考えられる。さらに特徴点という抽象化された表現を導入することにより、画像の構造に関する記憶との関連に迫ることが期待される。

今後は多重解像度の多層MRFモデルについて、その性質をさらに詳しく調べると共に、多種の情報との統合に関する検討も進めて行く。当面は、層間の相互作用の双方向化、および層間結合の学習アルゴリズムについて検討を進める予定である。

謝辞

有益な議論をしてくださったNHK放送技術研究所の三宅誠氏に深謝致します。また、研究の機会を与えてくださったATR視聴覚機構研究所淀川英司社長、認知機構研究室中根一成室長、そして日頃討論して下さるATR視聴覚機構研究所の皆様にご感謝いたします。

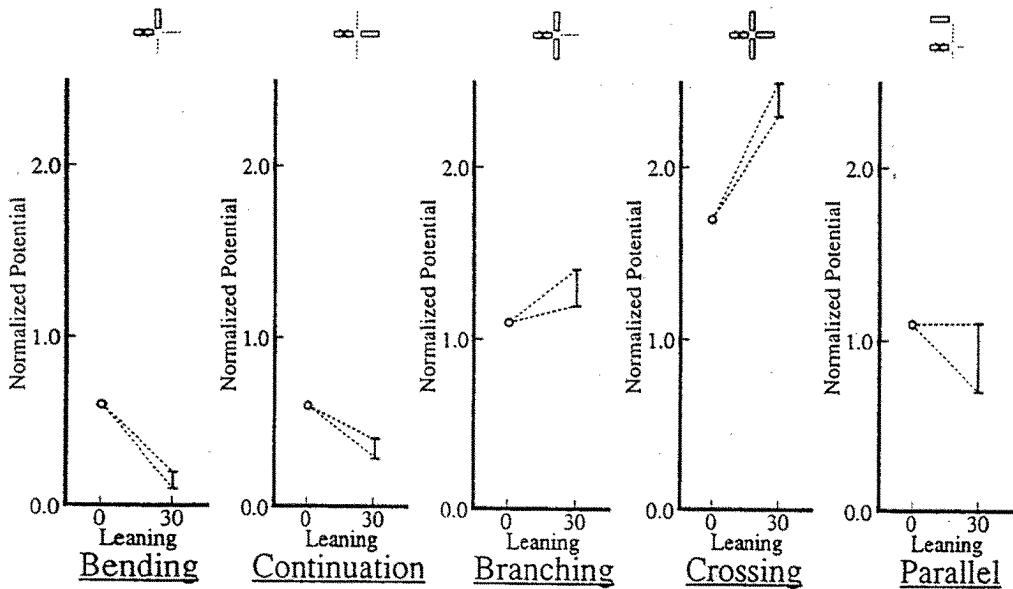


図3 3人の顔画像に対するエネルギー学習

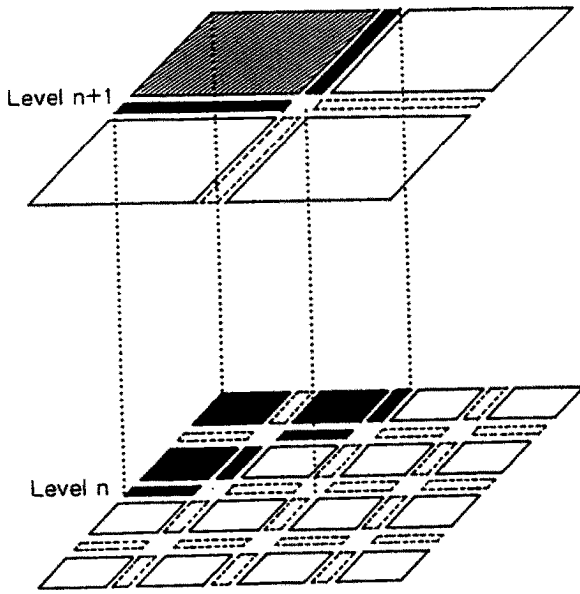
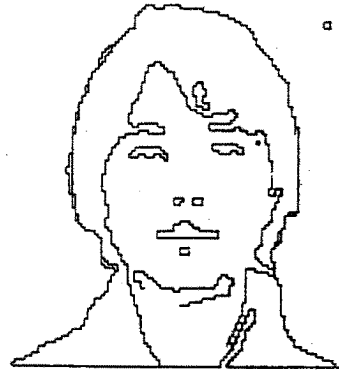
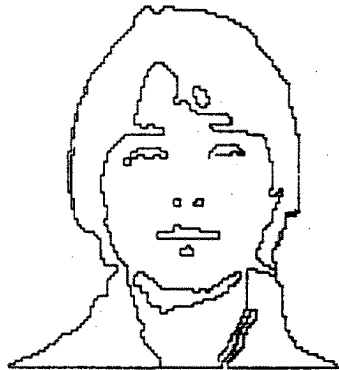


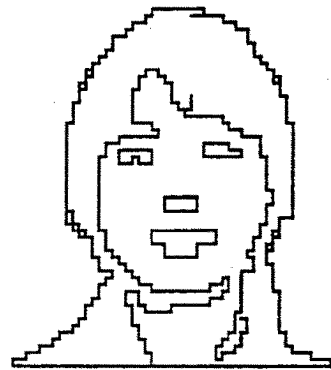
図4 多層MRFモデルの概念図



(a) 単層MRFモデルの抽出結果 (高解像度)



(b) 多層MRFモデルの抽出結果



(c) 単層MRFモデルの抽出結果 (低解像度)

図6 単層および多層MRFモデルによる抽出結果

[(b) は (c) の制約を加えたもの。]

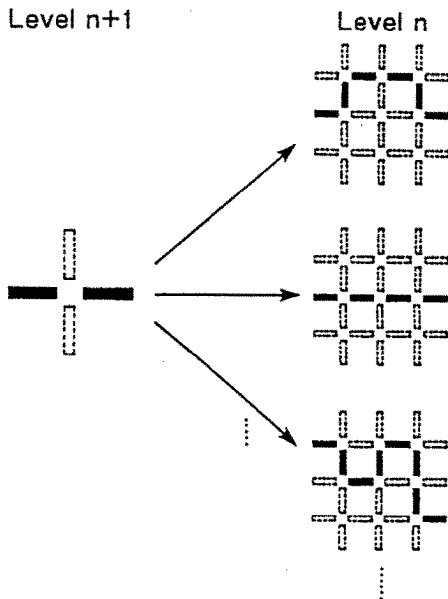


図5 階層間における線過程の配置の対応例

文献

- 1) D. Marr, "Vision", Freeman; 乾 敏郎, 安藤 広志訳: ビジョン—視覚の計算理論と脳内表現—, 産業図書, (1987).
- 2) T. Poggio, V. Torre, C. Koch, "Computational vision and regularization theory", Nature, Vol.317, pp.314-319, (1985).
- 3) S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images", IEEE Trans., Vol.PAMI-6, pp.721-741, 1984.
- 4) S. Hongo, M. Kawato, T. Inui, S. Miyake, "Contour extraction of images on parallel computer", IJCNN89 Proc., Vol.1, pp.161-168, 1989.
- 5) 本郷, 川人, 乾, 三宅, "エネルギー学習をする局所並列確率アルゴリズムを用いた輪郭線抽出", MBE88-181, pp.151-156, 1989.
- 6) 川人, 池田, 曾根原, 乾, 三宅, "画像情報処理と神経回路モデル", 人工知能学会誌, Vol.4, No.2, 1989.
- 7) H. R. Wilson, D. K. McFarlane, & G. C. Phillips, "Spatial frequency tuning of orientation selective units estimated oblique masking", Vision Research, Vol.23, pp.873-882, (1983).
- 8) A. B. Watson, "Summation of grating patches indicates many types of detector at one retinal location", Vision Research, Vol.21, pp.17-25, (1982).
- 9) T. Inui, "A model of human visual memory: data compression with multi resolution", SCIA-89 Proc., pp.325-332, (1989).
- 10) 乾, 三宅, "図形の構造記述と視覚記憶のモデル (I)", 電子情報通信学会研究会資料, MBE88-107, pp.91-98, (1989).