

仮説検証型の画像理解を想定した初期視覚のニューラルネットワークモデル

鶴田, 直之
九州大学システム情報科学研究院知能システム学部門

谷口, 倫一郎
九州大学システム情報科学研究院知能システム学部門

雨宮, 真人
九州大学システム情報科学研究院知能システム学部門

<https://doi.org/10.15017/5783>

出版情報：九州大学大学院総合理工学報告．12（3），pp.339-345，1990-12．九州大学大学院総合理工学
研究科
バージョン：
権利関係：

仮説検証型の画像理解を想定した初期視覚の ニューラルネットワークモデル

鶴田直之*・谷口倫一郎*・雨宮真人*

(平成2年8月31日 受理)

Neural Network Model for Early Vision

Naoyuki TSURUTA, Rin-ichiro TANIGUCHI, Makoto AMAMIYA

In image understanding, it is necessary to combine data-driven analysis and model-driven prediction adequately. Therefore recently, the approaches of hypothesis-verification style come to be active. In these approaches, there are two main difficulties: how to control the adaptation of hypothesis and how to speed up large hypothesis verification processes. For these difficulties, the approach based on neural networks, which are performed in asynchronous massive parallel, is seemed to be useful.

In early vision, there are few approaches from the above view point. In this paper, we, first, discuss that the model-driven prediction is also necessary for early vision. And, then, we propose a neural network model for early vision. The proposed network is constructed with three sub-networks based on Hopfield's network. The first one is for regularization with line process. The second one is for classification. The line process detect edges in the classified image, and is performed well on arbitrary scale in the regularization. The last sub-network enables higher level processes to control the scale locally. Finally, we present results of our experiments of edge detection based on the above method.

1. 緒 言

人工知能の分野では、人間の優れた視覚情報の処理機能の基本原理解を究明し、工学的に広く活用しようとする試みが古くからなされている。

我々が網膜あるいはカメラから入手する画像は冗長な情報を多く含んでいるため、入力画像をそのまま処理するという事は、システムを複雑にするばかりで経済的にも信頼性の点からみても現実的ではない。したがって、画像から外界を理解するためには、画像から適切な特徴を抽出して、その特徴を統合して抽象化するというデータ駆動型の解析が必要となる。

しかし、また一方では、画像は外界の様々な要因が二次元の濃淡情報に縮退して得られたものであるから、一般には画像から外界を推定することは、何らかの拘束条件を用いなければ解析的には解けない問題である。さらに、利用可能な拘束条件は、より一般的な世界を

対象とした場合に普遍的なものではなくなる。したがって、画像から外界を理解するためには、外界の状況や視覚情報処理の目的（例えば、本を探すという意志など）によって拘束条件（仮説）を使い分けるというモデル駆動型の予測あるいは注意の集中が必要となる。

データ駆動による解析とモデル駆動による予測や注意の集中を統合した画像理解システムをモデル化するために、従来から仮説検証型の試みがなされている。この試みでは、必然的に仮説の適用制御と仮説検証の高速化が問題となる。筆者らは、この問題は処理の実行順序を明示する手続き型のプログラミングでは効率よく記述し得ないものとし、画像理解を“画像を最もうまく理解するための仮説（パラメータ）の組み合わせを選択する組み合わせ最適化問題”と捉えた画像理解のためのニューラルネットワークモデルを検討している。

本稿では、従来の視覚システムモデルの問題点を明確にした上で、初期視覚と呼ばれる特徴抽出の第一段階においてもモデル駆動型の予測や注意の集中が必要であるという観点に立ち、ニューラルネットワークを用いて初期視覚システムをモデル化する。

*情報システム学専攻 職員

2. 視覚情報処理のモデル

2.1 従来の視覚システムのモデル

十分に成熟した人間は、その中枢に知識として構造化した世界を持っている。その世界は、世界モデルと呼ばれる。視覚情報処理とは、画像として得られた情報を抽象化かつ構造化して世界モデルと照合することによって外界を理解する一連の処理過程をいう。このような処理を実行する視覚システムのモデルは、アプローチの仕方によって様々であるが、一般に高次視覚、中間視覚、初期視覚の三つのサブシステムから構成される。

高次視覚は、抽象化かつ構造化した視覚情報を世界モデルと照合して外界を理解する確信の維持機能と、次の行動に必要な情報に注意を向けるために世界モデルをもとに外界の変化を予測する推論機能を果たす。世界モデルの記述例は、意味ネットワークや述語論理に基づいたものが多くみられる。最近では、脳の非同期の並列処理形態を陽にモデル化したコネクショニストモデル¹⁾がある。しかし、注意の集中など高次視覚の能動的な振舞いのモデル化はほとんど進んでいない。

中間視覚は、初期視覚によって得られた画像中の特徴を統合して構造化する役割を果たす。中間視覚の重要な機能は、画像に対応して二次元的に分布している特徴を統合して外界の三次元構造を復元する機能であるといわれている。中間視覚の初歩的なモデルは、パターン意識モデルにみられる。このモデルでは、処理が階層的に進み、下層から上層にいくにしたがって特徴は統合され抽象化される。

初期視覚は、外界を理解するために必要な最もプリミティブな特徴を画像から抽出する。初期視覚で抽出される特徴は、対象物の輪郭や領域、表面の曲率、対象物までの距離などで種類は多くはない。

2.2 従来のモデルの問題点

本節では、従来の視覚システムモデルの二つの問題点を指摘し、それぞれに対応してニューラルネットワークを用いる動機と従来の初期視覚モデルを改良する理由について論じる。

2.1節で述べたようなモデルにおいて、データ駆動型の解析では、初期視覚で入力された画像をボトムアップに処理して高次視覚で理解の結果を出力する。これに対して、モデル駆動型の予測では、高次視覚で発せられた要求をトップダウンに伝搬させて下位のサブ

システムを制御する。この二つの処理は、それぞれ独立したシステムではなく互いに有機的に結合したものであるから、一つのサブシステムを考えるとときに、その内部で行われる処理の実行順序は予め明示的には決定できない。したがって、視覚システムは、初期視覚あるいは高次視覚における状態の変化を伝搬させて出力を更新するような事象駆動型のシステムによってのみモデル化できる。この点から、非同期の並列処理形態にしたがって動作するニューラルネットワークによるモデル化は最も有力な方法の一つである。

次に、初期視覚は、画像を理解するために必要な特徴のプリミティブを抽出するものであった。ここで抽出されるプリミティブは、1節のはじめで述べたように必要な情報を失わない程度に十分に抽象化されたものでなければならない。この点から、初期視覚システムもトップダウン型の制御を受ける必要がある。しかし、トップダウン型の制御を考慮した初期視覚システムのモデル化はまだ行われていない。そこで、次節では、トップダウン型の制御に応じて任意の抽象度で特徴を抽出するためのニューラルネットワークによる初期視覚モデルを提案する。

3. 初期視覚のモデル化

3.1 従来の初期視覚モデル

初期視覚の役割の一つは、前節で述べたように雑音の除去と対象物の表面の曲率および表面までの距離などの抽出である。これらの問題は、画像の生成過程の逆変換を求める光逆問題として統一的に扱えることがPoggioらによって示された。これを標準正則化理論と呼ぶ²⁾。具体的には、画像データを $I(x, y)$ 、特徴（例えば、画像上の点 (x, y) に写っている対象物の可視表面までの距離）を $U(x, y)$ とすると、初期視覚の問題は、

$$A \cdot U = I \quad (1)$$

から $U(x, y)$ を推定する問題として定式化できる。ここで、 A は画像の生成過程を表す線形の演算子である。しかし、 U から I への変換過程では情報が縮退しているから、(1)式は何らかの拘束条件がなければ解けない。そこで、標準正則化理論では、安定化汎関数と呼ばれる合理的な拘束条件 $Q \cdot U = 0$ を付加して、次式

$$(1-\lambda)\|A \cdot U - I\|^2 + \lambda\|Q \cdot U\|^2 \quad (2)$$

を最小にする U を解として推定する。ここで、 λ は第一項と第二項を重み付けするパラメータでスケールと呼ばれる。また、 $\|\cdot\|$ は L_2 ノルムである。(2)式において A と Q は共に線形であるから、(2)式是最急降下法によって容易にかつ高速に解くことができる。標準正則化理論でよく用いられる安定化汎関数は、離散画素上で定義された画像に対して

$$\|Q \cdot U\|^2 = \sum_j \sum_i (\nabla U)^2 \quad (3)$$

で、滑らかな U の推定を要求する拘束条件である。(2)式にしたがって U を推定するネットワークモデルを Fig. 1 に示す。Fig. 1 のネットワークは、画素に対応したユニットを局所的に相互に結合したものである。

この標準正則化理論の問題点は、領域の輪郭線が考慮されていないために、輪郭線上で二つの領域を連続的に推定してしまう点である。これに対し、ライン過程と呼ばれる領域の輪郭線を検出するユニットを導入した神経回路網が提案された (Fig. 2)³¹⁴⁾。このモデルでは、輪郭線上にあるライン過程が二つのユニット

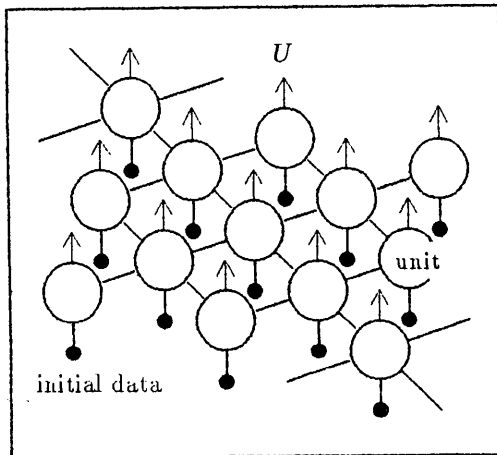


Fig. 1 Neural network model for early vision based on standard regularization theory, which is called Hopfield's network. In this figure, ● is initial data. And ○ is neuron unit whose value is $U(x, y)$.

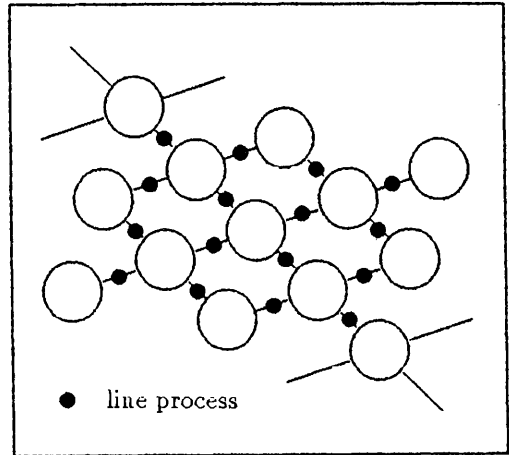


Fig. 2 Neural network model with line process for early vision. If edge is detected on a line process, the line process cut connection between the two adjacent units. In the figure, line processes are marked by ●.

間の作用を切断する。ライン過程の状態の計算の方法は、従来の輪郭線検出法を利用するものやマルコフ確率場モデルに基づいて最大事後確率推定を行うものなどがある。ライン過程を導入したモデルにおける U の推定は、最急降下法から

$$\begin{aligned} -\frac{dU(x, y)}{dt} = & (1-\lambda)(A \cdot U(x, y) - I(x, y)) \\ & + \lambda((U(x, y) - U(x+1, y))(1-V(x, y)) \\ & + (U(x, y) - U(x-1, y))(1-V(x-1, y)) \\ & + (U(x, y) - U(x, y+1))(1-H(x, y)) \\ & + (U(x, y) - U(x, y-1))(1-H(x, y-1))) \quad (4) \end{aligned}$$

である。ここで、 $V(x, y)$ と $H(x, y)$ がライン過程で、値が1のとき輪郭線を表す。以下では、初期視覚に対するモデル駆動型の制御の必要性の観点から、このモデルの問題点を明確にして二つの改良を加えたモデルを提案する。

3.2 任意の抽象度で領域を分割するモデル

前述のライン過程を導入したモデルは、モデル駆動型の制御が一切考慮されていない。具体的には、スケールをはじめとする各パラメータは、特定の画像群の統計的性質や雑音の特性から決定されるもので、これらの設定を変えると輪郭線が途切れたり不当に分岐

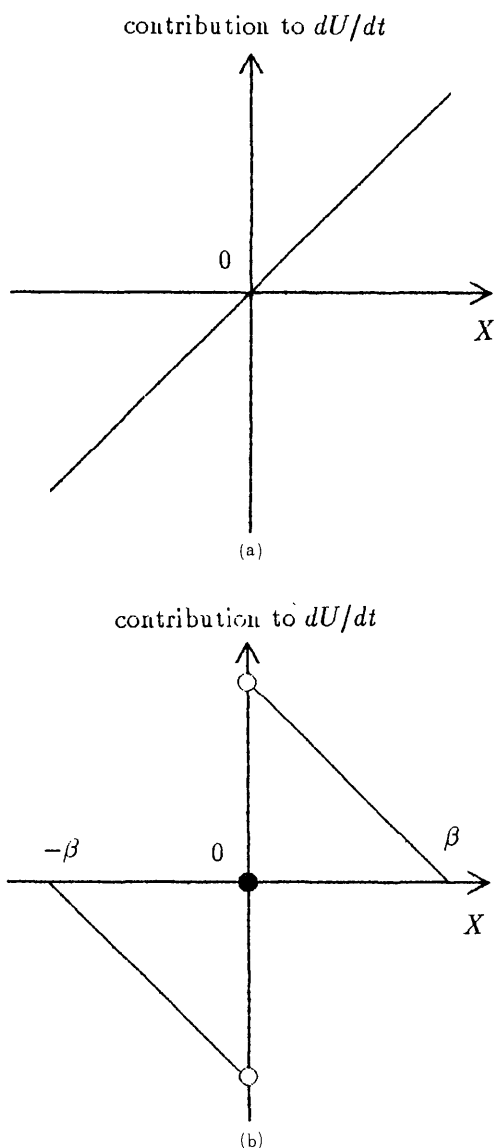


Fig. 3 The relation between difference of two adjacent units and its contribution to dU/dt . (see equation (4) and (5)). (a) regularization, (b) classification.

したりする⁵⁾。筆者らは、まずこの問題を解決し、任意のスケールで輪郭線が検出できるように群化処理を導入したモデルを提案する。

群化処理⁶⁾は、画像を濃度値の一樣な連結成分に分

割する領域分割の一手法である。群化処理後の画像では、同一領域内の画素には一樣な濃度値が与えられているため、濃度変化の有無を判定するだけで不当な分岐のない閉じた輪郭線が安定に抽出できる。ただし、検出する輪郭線の抽象度を制御することは困難である。このことから、群化処理と正則化処理は、互いにその欠点を補完し合う関係にあり、両者を併用することによって任意のスケールで輪郭線を検出できる。以下では、準備として群化処理のためのニューラルネットワークモデルについて述べる。

[準備]

群化処理のためのニューラルネットワークは標準正則化理論でモデル化されたネットワークとはほぼ同じで、データ U の時間変化則は、次式に従う。

$$-\frac{dU(x,y)}{dt} = G(U(x,y) - U(x+1,y)) + G(U(x,y) - U(x-1,y)) + G(U(x,y) - U(x,y+1)) + G(U(x,y) - U(x,y-1)) \quad (5-1)$$

$$G(z) = \begin{cases} \beta \operatorname{sgn}(z) - z & \text{if } -\beta < z < 0 \text{ or } 0 < z < \beta \\ 0 & \text{otherwise} \end{cases} \quad (5-2)$$

(5-1)式は、安定化汎関数による時間変化則 ((4)式の第二項以下) とよく似た形をしている。(4)式の第二項と(5-1)式の第一項の

$$X = U(x,y) - U(x+1,y)$$

に関するグラフをそれぞれ **Fig. 3(a), (b)** に示す。共に第一象限と第三象限を通る奇関数であることは、隣接する画素間で画素値の(重み付き)平均化が行われることを表している。これにより、群化処理によって画素の値が発散することはないのが分る。また、群化のグラフは傾きが負になっていることは、画素の値が近いものほどより近い値に更新されることを表しており、局所的に群化が行われることを表している。

群化処理後の輪郭線検出では、孤立雑音と共にコントラストの弱い微小領域の輪郭も消滅する。この消滅する微小領域の大きさとコントラストは、定数 β が大きくなるにつれて大きくなる。本手法では、定数の値を事前に適当な値に決定して固定する。すなわち、本

手法では、コントラストの弱い微小領域の一つ一つを抽出して解釈を与えることは無意味であるという前提を無条件に受け入れている。ただし、除去すべき微小領域の大きさやコントラストは正則化のスケールで制御するから、定数の値の決定はその下限（識別限界）を決定するだけのもので本質ではない。

[準備終り]

輪郭線検出の処理は、まず(4)式に基づいてデータが十分に安定するまで正則化した後、(5)式による群化処理を行い濃度変化の有無を調べて輪郭線を検出する。更に、検出した輪郭線の位置のみライン過程の値を1にして、以上の処理を繰り返す。

3.3 注意の集中を可能にするモデル

画像理解では、必要な情報だけを得るために目的にあったスケールを選択する必要がある、複数のスケールから同時に得られる情報をもとに画像を理解する試みがみられる⁷⁾。しかし、複雑な画像を対象とした場合、スケール間の対応付けの問題や資源の有効利用の点から、むしろ大きなスケールから初めて注意を引く領域に対してスケールを小さくするトップダウン型の理解が有効であると考えられる。注意を引く領域は、中間視覚の階層構造を逆向きにたどることとおおよその位置は指定できる。しかし、注意を向ける領域の広さや形を明確に指定できるとは限らないという問題が

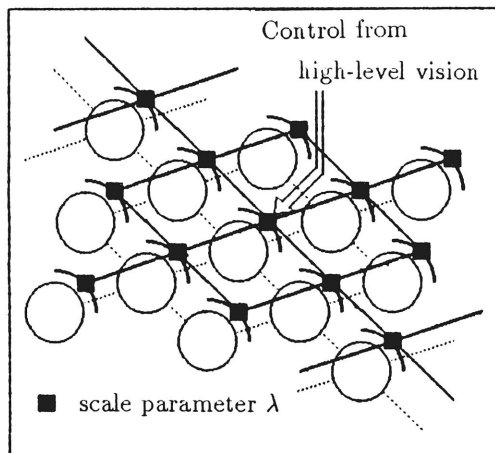
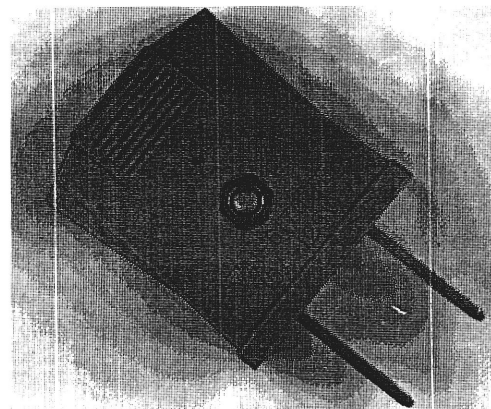
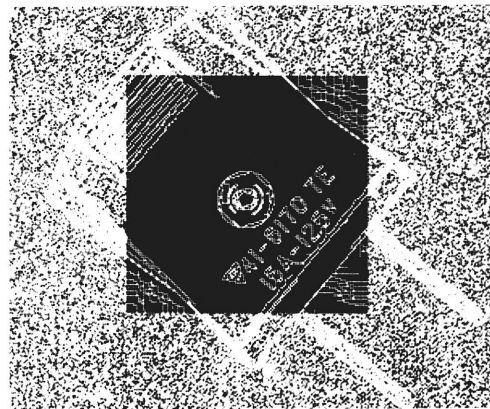


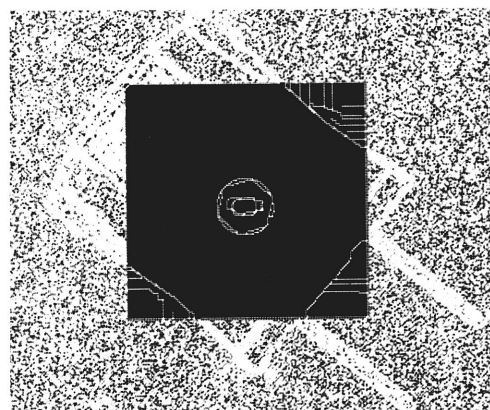
Fig. 4 Neural network model in which scale parameters are also constructed by Hopfield's network. This model enables to control the scale parameters locally. In this figure, each scale parameter is marked by ■: Line process are abridged in this figure.



(a)

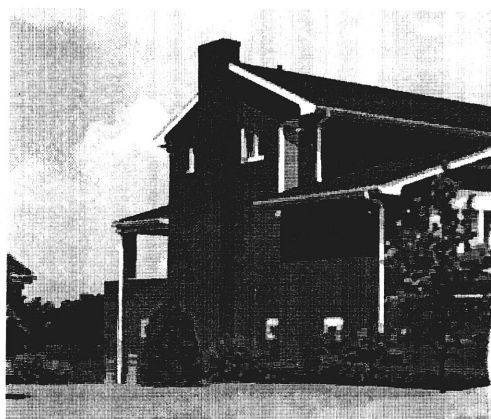


(b)

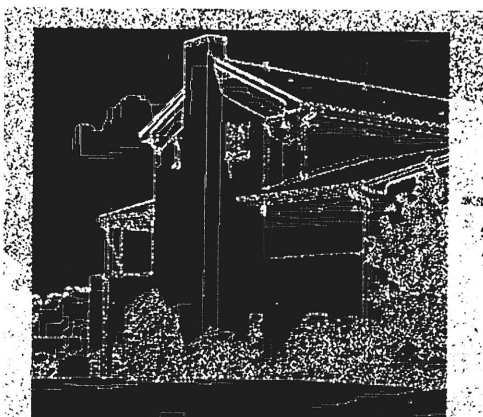


(c)

Fig. 5 (a) Original image: "Plug". (b) Detected edge after only classification process ($\beta = 17.0$). (c) Detected edge after regularization ($\lambda = 0.7$) and classification ($\beta = 17.0$).



(a)



(b)



(c)

Fig. 6 (a) Original image: "House". (b) Detected edge after only classification process ($\beta = 17.0$). (c) Detected edge after regularization ($\lambda = 0.7$) and classification ($\beta = 17.0$).

起こる。本稿では、この問題を解決するためにスケールを独立した相互結合型ネットワークで構成したモデルを提案する。これにより、注意の集中領域の中央でスケールを小さくするだけで、ネットワークの補完作用により制御の影響が伝搬してパラメータが滑らかに変化する。Fig. 4 にスケールをネットワークで構成したネットワークの構成図を示す。

4. 実験

本節では、3節のモデルの改良の正当性を実験によって確かめる。実験では、初期視覚の最も基本である雑音の除去問題に改良モデルを適用した。雑音の除去問題は、(2)式で A が単位写像の場合である。

4.1 任意のスケールでの輪郭線検出

2.2節で述べた手法を別々の撮像系から入力した“コンセント”と“ハウス”の画像に適用した実験結果を Fig. 5, 6 に示す。各図中の中央の正方形領域が処理の対象で、その周囲は未処理の領域である。256階調色の原画像(a)に対し、(b)は、群化処理のみ(スケール0.0に相当)を行い、濃度変化の有無を近似誤差を考慮して閾値1.0によって判定して得た輪郭線を示す。孤立雑音は除去され、連結性の強い輪郭線が検出されているのが分る。また、群化処理の定数 β は13.0に固定していることから、 β を固定しても実用的には問題ないことが確かめられた。(c)は、スケー

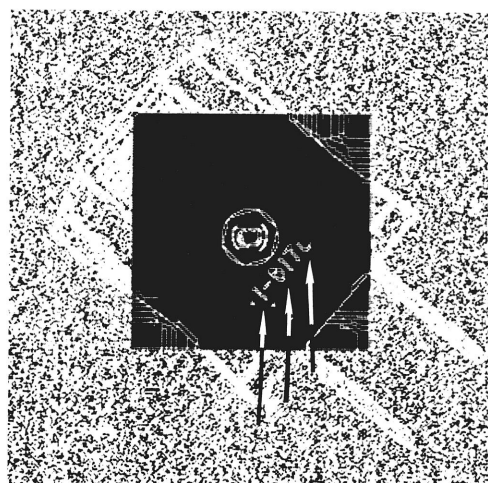


Fig. 7 Detected edge after scale-control for Fig. 5 (c). Controlled points is marked by three arrows.

ルを全域0.7と大きくして、(b)と同様の処理によって検出した輪郭線である。スケールの大きさに応じて連結性の強い輪郭線が安定に検出されているのが分る。

4.2 スケールの局所的制御後の輪郭線検出

スケールを局所的に制御した実験例を **Fig. 7** に示す。**Fig. 7** は **Fig. 5** の(c)の状態から図中の矢印で示した3点のみスケールを小さくした結果である。ネットワークの補完作用によって制御の影響が伝搬した領域のみ文字が認識できるようになった。また、スケールの局所的な変化によって不当な輪郭線が検出されていないことから、スケールが滑らかに変化しているのが分る。

4. ま と め

本稿では、仮説検証型の画像理解を想定し、任意のスケールで安定した輪郭線を抽出する初期視覚モデルを提案した。さらに、スケールパラメータを独立した相互結合型のネットワークで構成し、高レベルからの局所的な制御を可能にした。本稿で提案したネットワークは、奥行きデータの再構成問題など初期視覚の問題にそのまま利用できる。よって、今後は前述の

ネットワークを用いて画像理解のためのニューラルネットワークモデルを具体化してゆく。

参 考 文 献

- 1) J. A. Feldman, M. A. Fanty, N. H. Goddard: "Computing with Structured Neural Networks", *COMPUTER*, March, pp. 91-103 (1988).
- 2) T. Poggio, V. Torre and C. Koch: "Computational Vision and Regularization Theory", *Nature*, **317**, pp. 303-323 (1985).
- 3) S. Gemman and D. Gemman: "Stochastic Relaxation, Gibbs Distributions and Bayesian Restriction of Image", *IEEE Trans., PAMI-6*, pp. 721-741 (1984).
- 4) C. Koch, J. Marroquin and A. Yuille: "Analog" "Neural" Networks in Early Vision", *Proc. Natl. Acad. Sci. USA*, **83**, pp. 4263-4267 (1986).
- 5) 川人, 池田, 三宅: "神経回路網の学習と視覚情報処理", *テレビジョン学会誌*, Vol. 42, No. 9 (1988).
- 6) L. Davis and A. Rosenfeld: "Cooperating Processes for Low-level Vision: A Survey", *Artificial Intelligence*, **17**, pp. 245-263 (1981).
- 7) D. Marr, T. Poggio: "A Computational Theory of Human Stereo Vision", *Proc. R. Soc. Lond.*, **B204**, pp. 301-328.