研究論文

ハイブリッド画像理解システムにおける選択的注意モデルの
パラメータチューニング^{*1}三ツ森 芳幸^{*2}, 大 森 隆 司^{*3}メディアドライブ株式会社新技術開発研究所^{*2}, 北海道大学大学院情報科学研究科複合情報学専攻^{*3}

Parameter Tuning of the Selective Attention Model in a Hybrid Image Understanding System

Yoshiyuki Mitsumori^{*2} and Takashi Omori^{*3}Research & Development Center of Advanced Technologies, Media Drive Corporation^{*2}
Graduate School of Information Science and Technology, Hokkaido University^{*3}

Fukushima's Selective Attention Model is a model of biological vision system. It has engineering merit of deformation and position shift tolerant recognition and recollection of the recognized object in an input image. Based on this feature, we have proposed a method for hybrid image understanding in which each object is recognized, recollected and segmented sequentially even when the objects are overlapped and occluded. However, parameter setting for the fine object recollection is difficult with the Selective Attention Model and parameters search is necessary to tune the given image. In this paper, we propose a method for the parameter search based on an evaluation of image recollection precision in the hybrid image understanding process. A computer simulation result demonstrates the validity of the proposed method.

1. は じ め に

福島による選択的注意のモデル (Selective Attention Model : SAM) は, もととは生物の視覚系のモデルとして提案されたものであるが, 工学的な観点からも認識した物体の画像中での対応領域を入力図形の変形・位置ずれに対応しつつ抽出するという優れた性質がある¹⁾. 我々は, その性質を利用して背景のある画像から物体の領域を分離し, 順次的に画像を理解していく方法を提案してきた^{2, 3)}. 予め与えられた対象物体の位置関係などの背景知識から物体の誤認識の判定は可能であったが, SAM は入力によっては想起画像が不完全なことがあり, そのまま実画像への利用は難しかった. この問題に対して我々は, 記号処理と SAM を組み合わせたハイブリッド画像理解方式を提案し, 精度の高い画像想起を実現することで複数の物体が重なり合っ

た画像の順次的な理解が可能であることを示した^{4, 5)}. ところが実際の画像に適用するには SAM において特徴抽出の閾値パラメータを入力画像に合わせてチューニングする必要があった. 選択的注意モデルは階層的に局所的な特徴を抽出することで物体を認識し, さらに逆向きにたどることによって入力物体を想起することができるが, すべての局所特徴を正しく抽出できなければ認識はできても正しい想起画像は得られず, ひいては画像理解も成功しない. つまり認識と想起の両方に適した閾値を設定する必要がある.

そこで本稿ではハイブリッド画像理解方式の想起画像を用いて対象物体の認識・想起の精度を評価する手法によって SAM のパラメータチューニングが可能であることを示す. そのため, 想起画像の正確さを示す指標として, SAM とハイブリッド画像理解による想起画像の一致の度合い「一致率」と, 認識結果の一意性を示す指標「エントロピー」を定義する. そしてそれらをそれぞれ最大, 最小とするように山登り法によりパラメータ探索を行う. さらに本研究では, 画像のエッジ抽出パラメータにも同じ指標を適用し, より適

*1 2004年7月1日受付, 2005年1月18日受理

*2 〒 360-0037 埼玉県熊谷市筑波 3-195 番地 熊谷駅前ビル

*3 〒 060-1408 北海道札幌市北区北 14 条西 9 丁目

切なエッジ抽出パラメータの探索が可能であることも示す。それによって、知識駆動の画像理解の手法としてハイブリッド画像理解の方式が有効であることを実験的に確認した。

2. 選択的注意モデルのパラメータチューニング

2.1 選択的注意モデルのパラメータ

選択的注意モデル (SAM) には認識、想起を繰り返すことで自動的にパラメータをチューニングする機構が備わっているが、閾値検出感度を制御するパラメータや繰り返し回数を適切に設定しないと画像によっては二重にエッジを想起するなど所期の想定に反する特性があり、その動作制御は容易ではない⁴⁾。これまでに SAM の認識機構に相当する神経回路モデルであるネオコグニトロン⁵⁾の閾値 r については、谷川らが定性的な設定方法について述べ⁶⁾、さらに David らは実験的な設定方法について述べている⁷⁾が、撮影条件の違う個々の物体についてのチューニングなどの具体的な方法は述べていない。また、認識だけのモデルであるネオコグニトロンに最適なパラメータが必ずしも SAM の画像想起に最適であるとは限らない。物体の領域検出・重なり検出など精密な画像理解を行うには、個々の物体が認識できるだけでなく、その物体を構成する局所特徴すべてを背景から分離し正しく抽出しなければならない。

特徴を正しく抽出したことの検証には想起の機能が重要となる。ネオコグニトロンに想起の機能を付加したものが SAM であるが、ネオコグニトロン⁵⁾の「認識」に最適なパラメータが必ずしも SAM の画像の「想起」に最適であるとは限らない。また実画像の認識では対象物体が同じでも撮影条件などが変わると最適な閾値も変わり、パラメータを事前に決定しておくことは難しい。さらに画像中に複数の物体がある場合には、撮影条件によって最適なパラメータは個々の物体で変化する。よって認識だけではなく正しい想起画像が得られるように、対象とする画像や対象物体に応じて閾値を自動的に決定する必要がある。

この問題に対して我々は、パターン処理と記号処理の組み合わせによるハイブリッド画像理解システム⁵⁾が、記号知識によって比較的正確な想起画像を生成できることに注目し、それを SAM のパラメータチューニングに利用する。我々が使用するハイブリッドシステムは、SAM を用いたパターン処理部と記号処理部が共通の知識を持ち、それぞれの処理部が個別に画像想起を行うことが可能である。特に記号処理部による想起画像は入力画像の変形に対応しつつも、想起の欠

損がないという特性がある。したがって、二つの処理部の想起画像の比較によって SAM の想起の正確さを評価することで、パラメータの自動チューニングが可能となる。

2.2 システムの基本構造と動作

Fig. 1 に並列型ハイブリッド画像理解システムの構造を示す。ハイブリッド画像理解システムでは記号処理部とパターン処理部が並列に存在し、相互に補完することでより正確な画像理解を行う。記号処理とパターン処理はそれぞれ階層的な知識を持ち、また共通の知識を持つことで記号処理とパターン処理とが対応する各階層において相互補完が可能となっている。SAM は階層的な図形の局所特徴の抽出・学習、さらには逆引きの階層的な物体の想起が行えるため、本システムではパターン処理部に用いている。SAM が学習した局所特徴のコネクション情報を想起時に記号知識へと変換することで、記号処理とパターン処理に共通な知識を自動的に獲得する⁵⁾。

パラメータのチューニングは以下の過程により行う。

1. 入力画像にエッジ抽出フィルタ (閾値 T が必要) を適用しエッジ画像を得る
2. SAM により図形の認識を行う (SAM の特徴抽出のパラメータ r が必要)
3. 認識結果から SAM により想起画像 (以降、「パターン想起画像」と呼ぶ) を得る
4. SAM の認識細胞の発火パターンと SAM のコネクションより変換された記号知識からアフィン変換などを用い、入力図形の特徴の位置を推定し想起画像を (以降、「記号想起画像」と呼ぶ) を得る

Fig. 1 Structure of the hybrid image understanding system.

5. 「パターン想起画像」と「記号想起画像」を比較し、両者が一致するようにパラメータを調整する

ここで、SAM の特徴抽出のパラメータ r 及び前処理であるエッジ抽出フィルタの閾値 T の最適化が必要となる。

3. パラメータ探索の評価基準

3.1 想起画像の比較による探索

本手法では、入力画像の認識によって得られる SAM のアナログ的な計算処理による「パターン想起画像」と、ハイブリッド画像理解システムとしての記号処理的な計算過程による「記号想起画像」を利用する。我々はその両者を比較して、パターン想起画像と記号想起画像が一致するように山登り法でパラメータを調整し、結果として初期パラメータによるものより質のよい想起画像を得る。それには、二つの画像の一致度合いを測る基準が必要である。また同時に、認識出力の質についても同様の方式でパラメータの調整が必要であり、そのための基準も必要である。

また本システムは入力に線画像を用いているため、実画像を扱うにはエッジ抽出の前処理が必要である。エッジ抽出のパラメータもまた上記の二つの評価基準を用いることにより SAM の閾値と同様に探索することができる。

なお、画像理解では物体の認識結果はその場面において物体間の依存関係からある程度評価できる^{2,3)}。そこで本稿では物体の認識結果が正しいことを前提としたパラメータ探索を扱う。しかし後に述べるように、この制約はパラメータ探索の結果から緩和される。

3.2 一致率による想起画像の評価

二つの想起画像が同じであることを評価する基準として、「一致率」を次式のように定義する。

$$M = \frac{||N \& S||}{\max(||N||, ||LN||)} \quad (1)$$

N , S は対象図形のパターン想起画像、記号想起画像を表し、 LN は学習画像を入力したときの SAM のパターン想起画像を表す。 $||*||$ は発火している画素数、 $\&$ は画素単位の論理積を意味し、 $||N \& S||$ はパターン想起画像と記号想起画像とで一致した画素数となる。入力画像により図形を構成する画素数が異なるため規格化をするが、変形パターンが入力されることを考慮し、 $\max(||N||, ||LN||)$ で規格化を行った。したがって SAM と記号処理による想起画像が完全に一致すれば $M = 1.0$ となり、違いが大きいほど 0.0 に近づく。ここで入力物体に変形があり学習画像よりも小さい場

合、分母は $\max(||N||, ||LN||) = |LN|$ となり正しい想起画像が得られた場合でも M は 1.0 より小さくなる。この状態から M を大きくするようにチューニングを行うと分子側の値を大きく、つまり想起画像の画素数を多くするために余計な画素を想起する可能性がある。SAM は隣接する線を誤って想起することがあるが、記号処理による想起ではこのような誤想起は起こりにくい。そのため分子の値は大きくならず M の値も変化しないので余計な画素を想起するようなチューニングは行われない。

3.3 エントロピーによる認識結果の評価

SAM では、細胞の閾値パラメータ r を低く設定すれば特徴が抽出され易くなり、図形の特徴も想起され易くなる。一方で複数の認識結果が得られることが多くなり、誤認識が起こりやすくなる。画像の認識機構でもある SAM としては、一つの認識出力のみが明確に高くなることが望ましい。そこで複数の認識細胞が発火しないことの評価基準として、それぞれの認識出力細胞の値 P_i ($i = 1 \dots n$, n は認識出力細胞の数) を確率に相当する数値とみなして「エントロピー」を次式のように定義した。

$$E = \frac{-\left(\sum_i \frac{P_i}{\sum_j P_j} \log \frac{P_i}{\sum_j P_j}\right)}{\log n} \quad (2)$$

認識結果が一つだけであればエントロピー E は最低値 0 となるので、 E が低くなるようにパラメータ探索を行う。

3.4 探索方法

パラメータ探索は前記の二つの評価基準に対して山登り法を使用し、パラメータの更新方法には個々のパラメータを順番に変更していく逐次方式と、それらを同時に変えながら最適な変更の方向を探る同時方式の二通りの方法を検討した。また、SAM では $\frac{r}{1+r}$ が細胞の閾値になるので k 層目 ($k = 1, 2, \dots$) の細胞の閾値を

$$R_k = \frac{r_k}{1 + r_k} \quad (3)$$

とにおいて R_k を更新した。

逐次方式では R_k を $\pm \Delta R_k$ 変化させ評価値が大きくなる方へ R_k を更新する。この操作を下位層から上位層へと順番に繰り返し行う。一巡しても評価値が変わらない場合は変化量を $\pm 2\Delta R_k$ とし、同様に m 順しても評価値が変わらない場合の変化量を $\pm(m+1)\Delta R_k$ として探索する。この方式は、各瞬間の探索の次元が制限されるため、計算回数は少ないが、結果が局所解

に陥る可能性がある。

同時方式では R_k を 0, $\pm\Delta R_k$ 変化させ全ての k についての組み合わせの中で最も評価値が大きくなるようにそれぞれの R_k を更新する。 m 回繰り返しても評価値が変わらない場合の変化量は 0, $\pm(m+1)\Delta R_k$ として近傍のパラメータ空間を探索する。この方式は逐次方式と比較して大量の計算量が必要となるが、より良い結果が得られる可能性もある。

4. 実験

4.1 単一画像での評価

本実験で用いた学習画像を Fig. 2 に示す。原画像からのエッジ抽出にはラプラシアンフィルタを用い、人手により適当な閾値で二値化した。また、以下の単一画像の実験においてパラメータの初期値は同一とした。収束条件は逐次方式、同時方式それぞれの評価値が四巡および四回変化しなかった場合とした。

4.1.1 評価値の妥当性

以後、Fig. 3~Fig. 7 および Fig. 10 では、パラメータの探索回数に対する指標の変化と、その途中段階でのパターン想起画像と記号想起画像の変化を示していく。なお、入力画像は学習画像とは別角度から撮影したものであり、Fig. 3 から Fig. 5 のエッジ抽出時の閾値は実験者が目視により物体のエッジのみが抽出できる値に設定した。チューニング終了時に二つの想起画像が学習画像にもっとも類似していれば、チューニングは成功したといえる。実験では認識・想起する図形は学習画像とは異なるため類似の判断には学習した局所特徴を正しく想起できているかを実験者が目視により判断した。

Fig. 3(a), Fig. 3(b) に一致率を評価基準としたときの逐次・同時方式によるパラメータ探索の結果を示す。二つの方式とも一致率が上がるに従い想起画像が学習画像に近付いており、一致率 M は学習画像に近い想起画像を得るための評価基準としては妥当であると考

えられる。両方式とも初期値は同じであるが最終的な一致率、 R_k は異なる。これはパラメータの更新方式の違いから異なる局所解に落ち込んだためと考えられる。一方、エントロピーは 0 になっておらず、これは複数の認識結果が出力されていることを意味する。

Fig. 4(a), Fig. 4(b) にエントロピーを評価基準としたときの逐次・同時方式によるパラメータ探索の結果を示す。探索が進むにつれエントロピーが 0 に近付き認識出力が一つに絞られていく。しかし、想起画像はほとんど変化せず最終的に得られた想起画像は Fig. 3 の一致率を評価基準とした場合の方がよい。

(a) Sequential update search

(b) Simultaneous update search

Fig. 2 Learned patterns for the single object experiment.

Fig. 3 Result of the parameter tuning with the coincident rate (M). The upper line of the images inside is the recollected images by the Selective Attention Model. The lower line of the images inside is the recollected images by the symbolic processing.

Fig. 4 Result of the parameter tuning with the entropy (E).

これはエントロピーは認識結果を一つに絞り込む評価基準としては妥当であるが、よりよい想起画像を得るための評価基準とはなっていないことを意味する。

そこで新しい評価値 EM を一致率とエントロピーの線型和として次式のように定義した。

$$EM = \alpha M + (1 - \alpha) * (1 - E) \quad (4)$$

α は 1 以下の正定数である。一致率が大きく、エントロピーが小さくなるほど EM は 1 に近づく。 α を大きくすると一致率を重視することになり良い想起画像が得られる可能性が高くなるが、誤認識により複数の認識結果が出力される傾向も強まる。 α を小さくした場合はこの逆の傾向が見られる。 α を 0.7 とした EM を評価値としたときの探索結果を Fig. 5(a), Fig. 5(b) に示す。 α を 0.7 としたのは複数の認識出力が得られた場合でも物体間の依存関係などの背景知識から誤認

Fig. 5 Result of parameter tuning with linear sum of the coincident rate and the entropy (EM).

識を判断が可能であること、本システムでは複数物体の認識に一度認識した物体を入力画像からマスクアウトをするためにより正確な想起画像を必要とすることから一致率を重視した。探索が進み EM が大きくなるにつれエントロピーは小さく、一致率は大きくなっている。これは認識結果が一つに絞り込まれ、同時に想起画像も学習画像に近付いていることを意味するので評価値として EM は妥当だと考えられる。

4.1.2 エッジ抽出を含めたチューニング

これまでの実験では原画像からのエッジ抽出の閾値を実験者が決定していたが、この閾値の自動探索を試みる。入力画像が同じであってもエッジ抽出フィルタの閾値 T の値によって得られるエッジ画像は学習時と

(a) Sequential update search

(b) Simultaneous update search

Fig. 6 Result of parameter tuning including the edge extraction. The evaluation value EM and the threshold value T is shown.

は異なる。したがって T の値により一致率 M も変化する。学習した物体と同じ特徴が抽出できることを目標とするなら、特徴抽出の閾値 R と同じように $\pm \Delta T$ 変化させて一致率が大きくなるように閾値 T を調節すればよく、(3) 式において $R_0 = T$ と置くことによって SAM の閾値パラメータと同等に扱える。閾値 T を含めたチューニング結果を Fig. 6(a), Fig. 6(b) に示す。エッジ抽出にはラプラシアンフィルタを用い、収束条件は逐次方式、同時方式それぞれの評価値が五巡および五回変化しなかった場合とした。探索が進むにつれてフィルタによるエッジ画像および想起画像が学習画像に近付いており、評価値 EM を基準とすることで閾値 T もチューニングできることがわかる。

4.1.3 誤認識物体に対するチューニング

ここまでは認識結果が正しいことを前提としてパラメータチューニングを行ってきたが、誤認識物体に対

Fig. 7 Tuning process of the parameters on a mis-recognized object. A case that an object (B) is mis-recognized as an object (A).

してのチューニングも行い、誤認識時の評価値の振る舞いを検討した。Fig. 7 に Fig. 2(B) を Fig. 2(A) と誤認識したとしてパラメータ探索を行った結果を示す。図より、誤認識した物体に対してチューニングを行っても一致率が大きくならず、誤認識があった場合でも一致率からそれが判定できることがわかった。

4.2 同一画像中の複数物体の認識

本システムでは複数物体に対して、一度認識した物体の想起画像をマスクアウトして次の物体を認識するというを順次行う。しかし、パラメータ探索を物体ごとに初期値から行うと時間がかかる。そこで最初の物体での探索結果を次の認識での初期値とする方式で複数物体の認識を行った。その処理のフローを Fig. 8 に示す。探索方法には探索範囲が広い同時方式を用い、評価値が五回連続で変化しない場合 ($N_{th} = 5$) を収束と判定し、そのときの評価値が $EM_{th} = 0.5$ 以下の場合を誤認識とした。

学習画像を Fig. 9 に、探索結果を Fig. 10 と Fig. 11 に示す。初期パラメータでは Fig. 9 の (B) と (C) の二つの物体が認識された。認識出力が高い物体 (C) についてチューニングを行い 10 回目で収束したが、評価値が閾値 EM_{th} より低いいため誤認識と判断しパラメータを初期状態に戻した。次に初期状態において認識出力が 2 番目に高い物体 (B) についてチューニングを行った。その結果、評価値が EM_{th} より高い値で収束したので正しい認識と判断し、この物体の想起画像をマスクアウトした。さらに認識を行うと初期状態では認識されていた物体 (C) は認識されず、物体 (A) が認識されるようになった。そこで、この物体 (A) に対して

Fig. 8 Flow of the recognition algorithm for multiple objects.

Fig. 9 Learned patterns for the real image recognition.

チューニングを行うと、評価値が EM_{th} より高い値で収束したため正しい認識と判断した。これは最初は奥にある物体は手前の物体で一部が隠されているため認識されなかったが、手前の物体に対してチューニングを行いマスクアウトした結果、奥の物体の認識に対するノイズが減少し、見えている部分だけからでも正しい認識と正しい想起画像が得られたと考えられる。正しい想起画像が得られればより正確な補完画像⁴⁾が得られるので、物体が重なっている領域においてどちらの物体の特徴が多く欠けているかを調べることで物体の前後関係もより正確に判定できる。

Fig. 10 Recognition result of multiple objects in a real image.

Fig. 11 Tuning history of multiple objects in the real image. See section 4.2 for detail.

5. 考 察

Fig. 3～Fig. 6 の逐次方式と同時方式の結果を比べるとわずかではあるが逐次方式の一致率 M と評価値 EM の値が低かった。これは逐次方式の探索範囲が狭いために局所解に落ち込み易いためと考えられる。一方で計算量は、パラメータの更新方式において 1 ステップで

の評価値の比較回数は逐次方式は2回であるのに対して同時方式ではエッジ抽出の閾値を含めない場合で80 ($= 3^4 - 1$) 回, 閾値を含めると242回 ($= 3^5 - 1$) 必要である。逐次方式では4ステップを1サイクルとすることを考えても10倍に, 閾値を含めた場合は5ステップを1サイクルとするので約24倍にもなる。単一画像でエッジ抽出を含めない場合に1ステップにかかる時間は約2.5秒 (WindowsNT4.0, Celeron 450 MHz) であったのでこの差は大きい。何らかの計算量圧縮の方法が必要である。

一致率を評価基準としたパラメータチューニングでは下位の層ほど閾値の変化が激しかった。これはSAMが階層構造を持つために, 下位層のパラメータの変更がより直接的に一致率に影響するためと考えられる。一方, エントロピーを評価基準とした場合には上位層, つまり評価基準となる出力側に近いほど閾値の変化が激しかった。この結果から一致率とエントロピーのどちらを重視するかによって探索空間を絞り込んだり, 無駄な組み合わせのパラメータを削除したりすることで探索が速くなると考えられる。

複数物体でのチューニングでは物体ごとに周辺の物体でのチューニングの結果を利用しつつ最適なパラメータを探索することができた。しかし, 初期パラメータは人間が決定しているのでこれも自動化する必要がある。従来の画像処理における画像からのエッジ抽出パラメータ推定法などの組み合わせの評価が必要である。

6. ま と め

本稿ではハイブリッド型選択的注意モデルの閾値パラメータの探索の評価基準として「一致率」と「エントロピー」を定義し, その線型和を評価基準とした山登り法によるパラメータ探索を行い, より正確な想起

画像が得られることを確認した。また, この評価基準を用いることで画像の前処理におけるエッジ抽出の閾値も探索できることを示した。さらに, 複数物体においてもそれぞれの物体に最適なパラメータを探索することができた。「一致率」という評価基準を定義できたのは本システムが並列型のハイブリッドシステムであるからであり, このようなハイブリッドシステムは画像理解に有効だと考えられる。今後の課題は本システムに最適な探索方法を見つけることと計算量の削減である。

参 考 文 献

- 1) 福島邦彦 (1989): 神経回路と情報処理. 東京: 朝倉書店
- 2) 辻 義憲, 三ツ森芳幸, 佐伯大輔, 大森隆司 (1996): 記号処理とパターン処理の相互操作による画像理解システム, 画像の認識・理解シンポジウム, No.2, pp.19-24
- 3) Mitsumori, Y., Saeki, D., Tsuji, Y., and Omori, T. (1996): Hybrid architecture for image understanding by mutual evaluation of symbol and pattern processing, Proc. of ICONIP96, pp.444-449
- 4) 三ツ森芳幸, 大森隆司 (1999): 記号処理とパターン処理のハイブリッドによる領域分割, 信学技報, NC98-76, pp.57-64
- 5) 三ツ森芳幸, 大森隆司 (2004): 選択的注意モデルの学習情報の記号化によるハイブリッド画像理解の実現, 日本神経回路学会誌, Vol.11, No.2, pp.47-55
- 6) 谷川昌司, 福島邦彦 (1993): ネオコグニトロンにおける特徴抽出細胞のしきい値設定法, 信学論 D-II, Vol.J76-D-II, No.10, pp.2215-2222
- 7) Lovell, D.R., Downs, T., and Tsoi, A.C. (1997): An evaluation of The Neocognitron, IEEE Trans. Neural Networks, Vol.8, No.5, pp.1090-1105