ニューラルネットワークを用いた重なり合う物体の位置認識*

塩谷 成敏^{*1}, 福田 敏男^{*2}, 柴田 崇徳^{*3}
佐々木 恭助^{*4}, 竹内 直和^{*4}, 木下 達之^{*4}

Recognition of Overlapping Targets Using Artificial Neural Networks

Shigetoshi SHIOTANI, Toshio FUKUDA, Takanori SHIBATA,
Kyoussuke SASAKI, Naokazu TAKEUCHI and Tatsuyuki KINOSHITA

Image processing has been used in robot vision as a tool to establish the position of targets. Many algorithms using the template matching and artificial neural networks have been proposed for recognition. However, these algorithms are ineffective in computing cost and recognition accuracy in the case of recognition of overlapping targets. Therefore, we propose an effective algorithm for the recognition of overlapping targets using artificial neural networks which have learned movements of the human eye and the template matching. The artificial neural networks extract target candidates faster than conventional algorithms which find the candidates by moving templates of basic patterns in order. Then the template matching recognizes whether the candidates are targets or not. We compared the proposed algorithm with the conventional template matching algorithm with respect to the performance time and the recognition accuracy in the recognition of overlapping plant cells. The experimental results showed that the proposed algorithm recognized them faster than conventional algorithms and with high recognition accuracy.

Key Words: Neural Network, Image Processing, Recognition

1. はじめに

現在, 計算機の性能の向上により, 画像処理がロボットの視覚センサとして用いられるようになった。このとき, 画像処理は, ターゲットの位置を認識するのに使われる。これまでに, 原画像を二値化し, 得られた二値画像内からターゲットを抽出するための方法が多く報告されている。二値画像内のターゲットが, 他の物体と接触したり重複していない場合は, ラベリング処理を用いてターゲットの候補を切りだし, その後それとターゲットの基本形状を比較することによってターゲットを認識できる。しかし, 二値画像内のターゲットが他の物体と接触したり重複している場合は, ラベリング処理を用いてターゲットの候補を切出せないため, ターゲットを認識できない。このため, 重複する物体からターゲットを認識する方法が, いくつか報告されている。

(1) 円形分離法⁽¹⁾⁽²⁾は, ターゲットが円形である

場合にしか適用できず, 任意の形をしたターゲットを認識できないため問題がある。

(2) テンプレートマッチング法⁽³⁾は, 二値画面内で順次テンプレート(ターゲットの基本パターン)を動かして, 二値画像とターゲットの基本形を照合することによってターゲットとその位置を認識できる。しかし, テンプレートを順次動かすので照合する回数が多くなり, 認識時間がかかるのが問題であった。

(3) ニューラルネットワーク(以下, NNを用いる)を用いた方法⁽⁴⁾は, 二値画面内で順次ウィンドウを動かして画像を切出し, ウィンドウ内の画像をNNに直接入力することにより, ウィンドウ内のターゲットを認識できる。しかし, ウィンドウを順次動かすので照合する回数が多くなり, 認識時間がかかるのが問題であった。このほか, ウィンドウ内の画像が大きくなるとNNが大きくなり, 計算量が膨大に大きくなるのも問題点であった。

一方, NNを用い, 効率よく画面内でウィンドウを動かしターゲットを認識する方法が提案されている⁽⁵⁾⁽⁶⁾。この方法では, 人が, すばやくターゲットの方へ視点を移動させてターゲットを認識する目の動き, すなわち衝動性眼球運動(サックード)をNNに学習させている。しかし, この方法は, 画面内に他の物体と

* 原稿受付 平成5年12月6日。

^{*1} 三菱重工業(株)高砂研究所 (〒676 高砂市荒井町)。

^{*2} 正員, 名古屋大学工学部 (〒464 名古屋市千種区不老町)。

^{*3} 機械技術研究所ロボット工学部 (〒305 つくば市並木1-2)。

^{*4} 三菱重工業(株)名古屋研究所 (〒453 名古屋市岩塚町)。

重なっていないターゲットが1個ある単純な画像にし
か適用できず、ターゲットが重なり合うような複雑な
画像の認識には適用できない問題があった。

本論文では、文献(6)のNNを複雑な画像に適用
できるように改良し、文献(3)のテンプレートマッチ
ング法よりも高速に認識できる方法を提案する。提案
する方法では、人の眼球運動を学習したNNが、画面
内でウィンドウを動かしてターゲットの候補を含む画
像を切出した後、テンプレートマッチング法を用いて
ウィンドウ内のターゲットの候補がターゲットである
かどうかを認識する。この際、NNは文献(6)の方法
とは異なり、画面中の複数の箇所からターゲットの候
補を探索するようにした。このほか、探索開始点から
のNNの探索終了条件をより複雑にした。以下に、終
了条件を追加する必要性を述べる。文献(6)の探索ア
ルゴリズムをそのまま用いると、NNはターゲットを
見つけるまでウィンドウを動かし続けるため、探索開
始点の周辺にターゲットがない場合では、画面の広範
囲を探索してしまう。NNが、他の探索開始点からも
同様に探索を行うと、同じ領域を何回も重複してター
ゲット候補を探すことになり、計算効率が非常に悪く
なる。このため、探索開始点からのNNの探索を終了
させる条件をより複雑にする必要がある。探索終了条
件については、2章で詳細に述べる。

提案する方法は、人の眼球運動を学習したNNを
用いてウィンドウを動かすため、単純にウィンドウを
順に動かすのに比べてターゲットを含む画像を効率よ
く抽出できる。また、NNは、ターゲット候補の位置
を大まかに抽出するだけなので、文献(4)のようにター
ゲットの大きさに応じた大きなNNを必要としない。
また、テンプレートマッチング法が誤認識しやすい画
像を、NNにより前処理的に排除できるため、誤
ったターゲットの抽出が少ない。

本論文では、提案する方法の有効性を示すために重
複したセロリの魚雷形不定胚⁽⁷⁾の認識実験を行った。
実験では、計算効率と認識精度に関して、文献(3)の
テンプレートマッチング法と比較して良好な結果を得
ることができた。

2. 重複したターゲットの認識

画像中のターゲットを認識する前処理として、二値
化処理がよく用いられる。しかし、最良の二値画像を
得ることは難しく、ターゲットが他の物体と接触した
形で抽出されることが多い。図1は、顕微鏡下でCCD
カメラを通して取込んだセロリ細胞の像である。図2
は、分散比を評価関数とする自動しきい値法⁽⁸⁾で二値

化した像である。このように細胞が重複する場合、1
章で述べたように、個々の細胞を認識することが難し
い。

本論文では、図3に示すような流れで重なり合うター
ゲットの位置を認識する。図4に示すように、NNは、
ある探索開始点からウィンドウを移動させて、ターゲ
ットの候補を大まかに切り出す。この時、NNは、「ター
ゲット候補を見つける(図3中のstop状態)」か「ウ
ィンドウを動かす方向がわからなくなる(図3中の
fault状態)」まで、ウィンドウを動かす。fault状態の
場合、提案する方法は、テンプレートマッチングを行
わずに、NNを用いて別の探索開始点からターゲット
の候補を探す。stop状態の場合、図5に示すように、
複数の基本パターンをウィンドウ内で移動させながら
テンプレートマッチングを行い、ウィンドウ内の画像
と最も一致する基本パターンとその位置を探す。提案
する方法では、輪郭線と面積の照合率がしきい値を超
えると、それをターゲットと認識し、その位置をター
ゲットの位置として認識する。そして、再び別の探索

図1 原画像

図2 二値画像

図 3 認識のながれ

図 4 ニューラルネットワークを用いたターゲット候補の抽出

開始点から NN にターゲットの候補を探索させる。また、ウィンドウの動きが単振動のような場合、fault 状態と同様に処理する。fault 状態、stop 状態の詳細は、3 章で述べる。NN がすべての探索開始点から探索し終えるまで、この処理は繰返される。

図 6 は、NN の探索開始点の配置図である。探索開始点の数は、ウィンドウの大きさ (W) と探索点間の距離 (L) によって定まる。 L の値は、画像中に物体が密集している (二値化画像中の黒点部分が多い) 場合は小さく、そうでない場合は大きめに設定する。また、 W はターゲットの大きさよりも少し大きく設定する。

図 5 テンプレートマッチングによるターゲットの認識

図 6 ニューラルネットワークの探索開始点の位置

このように、複数の始点から探索させると同じターゲットを複数抽出することがある。このため、本論文では、抽出されたターゲットで近距離にあるものは、同一のものとみなし、それらの中から最も基本パターンと照合するものを選ぶ。ここで、面積の照合率と輪郭線の照合率の和を評価に用いた。ターゲット間の距離には、ユークリッド距離を用い、距離が 50 画素以内ならば同一のターゲットであるとした。

3. ニューラルネットワークを用いたターゲット候補の抽出

3.1 ニューラルネットワークの構成 図 7 に、NN の構成図を示す。NN には、入力数 100 個、出力数 10 個の三層の多層パーセプトロン⁽⁹⁾を用いた。NN の入力として、二値画像中を動くウィンドウ内の画像を 10×10 に圧縮した画像を用いる。

NN の入力は、以下の式で示される。

$$N(m, n) = F(m, n) \left\{ \sum_{Y=P_Y+\frac{mW}{10}}^{P_Y+\frac{(m+1)W}{10}-1} \sum_{X=P_X+\frac{nW}{10}}^{P_X+\frac{(m+1)W}{10}-1} V(X, Y) \right\} \times \left(\frac{W}{10} \right)^2 \quad (1)$$

ここで、

$N(m, n)$: ニューラルネットワークの入力 ($1 \leq m, n \leq 10$)

P_X, P_Y : ウィンドウの位置 ($1 \leq X, Y \leq 256$)

図7 ニューラルネットワークの構成図

W : ウィンドウの大きさ ($W=10, 20, \dots$)

$V(X, Y)$: 位置 (X, Y) における二値画像の画素値

$F(m, n)$: NN のフィルタ

である。

NN のフィルタは、0 と 1 の値をとり、認識するターゲットの形に応じてフィルターを決定する。

NN の各層の入出力関係は、以下の式で表される。

$$I_i = N(m, n) \quad (i=10(m-1)+n) \dots\dots\dots (2)$$

$$H_j = f(\sum_j V_{ji} I_i + C_j) \dots\dots\dots (3)$$

$$O_k = f(\sum_j W_{kj} H_j + D_k) \dots\dots\dots (4)$$

$$f(X) = \frac{1}{1+e^{-X}} \dots\dots\dots (5)$$

ここで、 I_i, H_j, O_k は、それぞれ i 番めの入力、 j 番めの中間層ニューロンの出力、 k 番めの出力層ニューロンの出力である。また、 V_{ji}, W_{kj}, C_j, D_k はそれぞれ、 j 番めの中間層ニューロンと k 番めの出力層ニューロンの結合荷重、 j 番めの中間層ニューロンと k 番めの出力層ニューロンのオフセット値である。 $f(x)$ は、シグモイド関数である。

そして、NN の出力は、ウィンドウを動かす方向である。それらは、上、下、左、右、左上、左下、右上、右下、停止(stop)、失敗(fault)の10種類である。ある画像が入力されると、NN は、出力層内で最大値をもつニューロンを選び、そのニューロンに対応する移動方向を出力する。

NN は、入力画像中の複数の始点の一つから探索を開始し、「停止」あるいは「失敗」と認識するまで、ウィンドウを動かしてターゲットを探索し続ける。ここで、「停止」はターゲット候補を抽出したことを意味し、「失敗」は移動方向を判定できなくなったことを意味する。

3.2 ニューラルネットワークの学習 図8に教師パターン例を示す。NN の教師信号として、ウィンドウを動かす方向を与える。教師信号は、0 と 1 で記

図8 ニューラルネットワークの学習パターン例

述されている。例えば、教師が「上」ならば、教師信号として「1 000 000 000」が出力層のニューロンにそれぞれ与えられる。教師は、経験に基づいて以下の判断基準により与える。移動方向に関しては、ウィンドウから見た物体のある方向、「停止」はウィンドウ内にターゲット全体が含まれる画像、「失敗」はウィンドウ内に物体がない場合（ウィンドウ内に白い部分が多い場合）の画像やターゲット以外の物体の画像を用いている。

NN の学習には、バックプロパゲーション法を用い、教師信号との誤差を減らすよう行う。以下に、教師信号との誤差、学習式を示す。

$$E = \sum_{p=1}^P \sum_{k=1}^{10} (T_{pk} - O_{pk})^2 \dots\dots\dots (6)$$

$$V_{ji}(t+1) = V_{ji}(t) - A \frac{\partial E}{\partial V_{ji}} \dots\dots\dots (7)$$

$$W_{kj}(t+1) = W_{kj}(t) - A \frac{\partial E}{\partial W_{kj}} \dots\dots\dots (8)$$

$$C_j(t+1) = C_j(t) - A \frac{\partial E}{\partial C_j} \dots\dots\dots (9)$$

$$D_k(t+1) = D_k(t) - A \frac{\partial E}{\partial D_k} \dots\dots\dots (10)$$

ここで、 $E, P, T_{pk}, O_{pk}, t, A$ は、それぞれ教師との誤差、学習パターン数、 P 番めの学習パターンに対する k 番めの出力層ニューロンの目標値と出力値、学習回数、学習係数である。

4. テンプレートマッチングによるターゲットの認識

NN が、「停止」と認識して、ターゲット候補を含む画像を切り出したとき、その画像内で複数の基本パターンを順に移動させてテンプレートマッチングを行い、基本パターンの輪郭線と面積(領域)が最も一致する基本パターンとその位置を探す。以下に、本論文に用いたテンプレートマッチングの手順を示す。

(1) ウィンドウ内の輪郭線像を求める [図5(b)]。

(2) (1)の輪郭線を3回4近傍に膨張した輪郭線像を作成する [図5(c)]。ターゲット候補と基本パターンの間に多少の形の相違があっても認識できるように、膨張した輪郭線を用いる。

(3) ウィンドウ内で、複数の基本パターンを順に3画素ずつ移動させながら(2)の輪郭線像と照合していく[図5(d)].そして、最も輪郭線の照合率 M_c が大きくなる基本パターンとその位置を求める[図5(e)].照合率は、以下の式で表せる。

$$M_c = N_m / N_b \quad \dots\dots\dots (11)$$

ここで、 N_m, N_b は、それぞれ基本パターンと(2)の輪郭線像で一致した画素数、基本パターンの輪郭線画素数である。

(4) (3)の照合率がしきい値 T_c を越えた時、基本パターンとウィンドウ内の画像間の面積照合率 M_a を求める[図5(f)].照合率は、以下の式で表せる。

$$M_a = A_m / A_b \quad \dots\dots\dots (12)$$

ここで、 A_m, A_b は、それぞれ基本パターンと(1)の二値像で一致した面積、基本パターンの面積である。

そして、照合に用いた複数の基本パターンの中から輪郭線の照合率と面積の照合率の和が最大となる基本パターンを選び、そのパターンが面積照合のしきい値 T_a を越えた場合、ターゲット候補をターゲットとして認識し、最も照合した位置をターゲットの位置として認識する。

(5) 複数のターゲットが抽出された場合、近距離にあるターゲットは同一のものとみなし、その中で最も基本パターンと照合したものを選ぶ(2章に詳細を記した)。

5. 認識実験と結果

提案する方法が、総当たりのテンプレートマッチン

図9 本実験で用いた四つのニューラルネットワーク

グ法(画像内でターゲットの基本パターンを順に動かしてテンプレートマッチングを行う方法)と比較して、計算効率と認識精度に関して効率よく重複したターゲットを認識できることを示すために、セロリの魚雷形不定胚の認識実験を行った。実験では、NNが学習に用いなかった20枚の画像を用いた。セロリの画像は、顕微鏡を通して倍率4倍で取込んだものを用いた。また、照明の強さは、一定に固定した。認識する魚雷形不定胚は、すべて他の細胞と接触している。まず、認識の前処理として、自動しきい値法⁽⁸⁾を用い二値化した。そして、両者の方法により、二値画像中の魚雷形不定胚の位置を認識した。

ところで、魚雷形不定胚は、細長い形をしているため、回転により形が変化する。NNは、回転したパターンを認識できないため、本論文では、図9に示すような4種類のフィルタと、それに対応する四つのNNを用いて回転した魚雷形不定胚の候補を探索した。

四つのNNは、それぞれ縦、横、左斜め、右斜めに向いた魚雷形不定胚の位置を大まかに認識する。図9中フィルタの黒い部分(値は1である)は、NNの可視範囲を示す。フィルタの形は、縦、横、左斜め、右斜めに回転したターゲットを含む形にした。実験に用いた四つのNNは、式(1)に示すように各フィルタを通

表1 ニューラルネットワークのニューロン数と基本パターン数

Neural network No.	1	2	3	4
Target				
Number of Neurons	Input layer	100	100	100
	Hidden layer	15	15	15
	Output layer	10	10	10
Number of training data	52	67	69	67
Number of basic patterns	5	6	5	7

図10 本実験に用いた基本パターン

したウィンドウ内の正規化画像が入力される。四つの NN は、3 章で述べた判断基準に基づいて経験的に決定された各 NN に対応する学習データを学習している。ただし、各 NN のターゲットは、それぞれ縦、横、右斜め、左斜めに回転した魚雷形不定胚とした。つまり、図 9 で縦に向いた魚雷形不定胚を探す NN 1 の場

(a) 認識過程 1

(b) 認識過程 2

(c) 認識過程 3

(d) 認識過程 4

図 11

合、横や斜めに向いた魚雷形不定胚の画像はターゲットでないため、教師データとして「失敗」を与えた。実験に用いた NN のニューロン数とその学習データ数、テンプレートマッチングに用いた基本パターン数を表 1 に示した。図 10 に、本実験で用いた基本パターンを示した。四つの NN は、それぞれ図 6 に示すような位置から魚雷形不定胚の候補を探索した。ここで、 $W=70$, $L=25$ に固定した。この場合、探索開始点の総数は 64 個である。また、NN が 1 回の処理で動かすウィンドウの移動量は、三画素とした。

そして、NN の処理後に行うテンプレートマッチングは、輪郭線と面積の照合率のしきい値を $Tc=0.8$, $Ta=0.8$ に設定して行った。図 11 に、提案した方法を用いた魚雷形不定胚の位置を認識する過程を示した。

一方、総当たりのテンプレートマッチング法では、ウィンドウにより画像を切り出さず、直接二値画像の左上から順に複数個の基本パターンを動かして、提

図 12 ウィンドウの移動回数

図 13 ターゲット数とターゲット候補数

図 14 テンプレートマッチング回数と認識時間

図 15 認識された細胞の数

案する方法で用いたテンプレートマッチングと同じ方法により行った。

図 12 は、20 枚の画像の認識において、四つの NN のウィンドウの移動回数と文献 (4) のように単純にウィンドウを順に動かした場合の移動回数を示している。提案する NN の場合、一画素あたり平均 553 回に対して、単純にウィンドウを順に動かす場合は、3 844 回必要だった。このことから、NN を用いた方が、単純に順にウィンドウを動かすより効率よくターゲット候補を含む画像を切り出せることがわかる。

図 13 に、四つの NN が抽出した魚雷形不定胚の候補の数とターゲットの数を示した。抽出した候補の数は、一画面あたり平均 28 個であった。本論文で用いた方法は、多点から探索を行うため、実際のターゲット数よりも抽出した候補の数が多くなった。また、20 枚の画像中の魚雷形不定胚は、四つの NN により、すべてターゲット候補として抽出された。提案した方法では、NN が切り出したターゲット候補を含むウィンドウ内の画像に対してテンプレートマッチングを行い、ターゲットを認識している。このため、NN が抽出したターゲット候補の数は、テンプレートマッチングを行う画像数となる。

一方、総当たりのテンプレートマッチング法では、ウィンドウを動かして画像を切り出さずに、直接図 10 に示した基本パターンを順に画面内で動かして、テンプレートマッチングを行い、ターゲットを認識する。テンプレートマッチングは、提案する方法の中で用いたテンプレートマッチングと同じ方法で行った。

図 14 に、両者の方法のテンプレートマッチングによる照合回数と認識時間の比を示した。総当たりのテンプレートマッチング法の照合回数は、提案した方法より平均 12 倍多かった。これまで、NN の計算量 (NN のウィンドウ移動回数) と両者のテンプレート

図 16 誤認識された細胞の数

マッチングの計算量(照合回数)を示したが、両者の計算効率を比較するには、二つの計算量を合せて考慮する必要がある。このため、両者の方法の認識時間を比較した。図 14 に示すように、総当たりのテンプレートマッチング法は、提案した方法の平均 5 倍の認識時間が必要だった。このことから、計算効率において、提案した方法は総当たりのテンプレートマッチング法よりも優れていることがいえた。本実験において、テンプレートマッチングに用いた基本パターンの総数は、23 個であったが、この数が増えることによって認識時間の格差が広がると思われる。

次に、認識率に関する結果を図 15 と 16 に示した。提案した方法の平均認識率は、総当たりのテンプレートマッチング法と同じ 93% であった (20 枚の画像の中にあつた 28 個の細胞の内、26 個を認識した)。認識率が同じになったのは、提案した方法が最終的にはテンプレートマッチング法を用いて、ターゲットを認識したためである。一方、誤認識した数に関しては、総当たりのテンプレートマッチング法が 12 個に対して、提案した方法では 3 個であった。これは、NN がターゲットの候補を抽出する際、テンプレートマッ

ング法がターゲットと間違えやすい画像を取り除いているためである。この結果から、提案する方法は誤認識する数が少ないため、総当たりのテンプレートマッチング法と比較して認識効率がよいといえる。

6. ま と め

本論文では、ターゲットが重なり合うような複雑な画像にいてもターゲットの位置を効率よく認識できる方法を提案した。そして、提案した方法を用いて重複したセロリの魚雷形不定胚の認識を行い、一般的によく用いられる総当たりのテンプレートマッチング法と比較した。その結果、提案した方法は、総当たりのテンプレートマッチング法と同精度で、誤認識することが少なく短時間に魚雷形不定胚を認識できることを示した。提案した方法は、認識の前処理として画像を二値化し、二値画像中のターゲットの認識を行ったが、今後原画像中のターゲットが直接認識できるように発展させていく予定である。

文 献

- (1) 木瀬・ほか2名, 組み合わせ的手法による重なり合った粒子像の画像処理, システム制御情報学会論文誌, 2-11 (1989), 363-369.
- (2) 福田・ほか5名, 画像処理を用いた生物細胞の認識(第1報, ニューラルネットワークを用いた植物細胞の生死認識), 機論, 57-542, C (1991), 3189-3196.
- (3) 尾崎・ほか1名, 画像処理-その基礎から応用まで (第2版), (1983), 241, 共立出版.
- (4) 阿居院・ほか2名, 神経回路網を用いた白黒集合写真からの顔領域の抽出, 電子情報通信学会春季全国大会予稿集, (1992), 7-268.
- (5) 今井・ほか2名, 3層BPモデルを用いたノイズのある原画像に対する手書き文字認識, 電子情報通信学会論文誌, D-II (1992), 314-323.
- (6) 鷲澤, サッケード機構のための視覚神経網, 電子情報通信学会論文誌, D-II (1990), 633-640.
- (7) 小林・ほか1名, 人工種子大量調製方法の開発, ケミカルエンジニアリング, (1992), 18-21.
- (8) 大津, 判定および最小2乗基準に基づく自動しきい値選定法, 電子情報通信学会論文誌, D (1979), 349.
- (9) Rumelhart, D. E., A Learning Internal Representation by Error Propagation, *Parallel Distributed Processing*, (1986), 318-362, MIT Press, Cambridge, Mass.