

Second-Order Neural Network を用いた移動物体領域の抽出

茶志川孝和(*,**), 武藤 佳恭(***)

動画から移動物体を抽出する方法として、従来から差分法が使われている。しかし、差分法による抽出では、移動物体の全体が得られないという問題がある。その為、モルフォロジフィルタによるノイズ除去と欠損個所の補完を後処理として行う方法が知られている。ところが、撮影環境によっては後処理が有効に機能しない場合も多く、また演算コストの高さも指摘されている。本研究では、フレーム間差分情報を外部入力として動作する Second-order Neural Network を用いた移動物体領域抽出手法を提案する。提案手法と従来手法を用いて、人工動画による実験と実動画によるシミュレーションを行った。それらの結果、提案手法は従来手法に比べノイズに影響されにくく、また抽出領域の欠損が少ないことが示された。また、実行時間に関しても従来手法に比べ約 30%短縮されることが確認された。

Extracting moving object areas based on Second-Order Neural Network

Takakazu Chashikawa(*,**) and Yoshiyasu Takefuji(***)

The frame difference technique has been widely used for extracting moving objects. However, the technique can not extract the whole shape of the moving object. A morphology filtering is used for reducing noises and complementing lacks of extracted objects as an aftertreatment to deal with the problem. The morphology filtering is not always available under various conditions and its calculation cost is high. This paper proposes a new method to extract moving object areas, based on the frame difference technique and the Second-order Neural Network. We experimented a proposed method and a conventional method using artificial movies and real movies. The experimental result shows that the proposed method has a better performance for a noise reduction and for an extraction of whole moving objects than the conventional method. The execution time of the proposed method is about 30% less than the conventional method.

はじめに

コンピュータビジョン研究において、動画からの移動物体の抽出は基本的かつ非常に困難な問題の一つである。移動物体を抽出するもっとも直接的な手法として、差分画像を利用した方法がある。差分画像を利用した方法は、計算が容易でハードウェア化がしやすいなどの利点がある。しかし、照明環境の変動がノイズとして出力される、移動物体の全体が抽出されないことがあるなどの問題もある。そのため、モルフォロジ演算によるフィルタ処理を適用する方法が広く知られている。しかし、撮影環境によってはうまく機能しない、計算コストが高いなどの問題が指摘されている。一方、Second-order Neural Network(以下、SONN)を用いて移動物体を抽出する研究として、Feed Forward Shunting(以下、FFS)が提案されている[1]。FFS は、空間的に近傍との結合と入力に関係付けることで、ノイズの影響を受けることなく時系列に動く矩形波の進行方向の先端エッジを抽出することを可能としている。しかし、扱っている空間が一次元であることや、検出できる方向が決められているなど、一般の動画処理に応用するには問題がある。また、画像処理への応用研究が多いニューラルネットワークとして Pulse Coupled Neural Network(以下、PCNN)が知られている。PCNN は、Eckhorn らが猫の視覚野におけるニューロン間の同期現象を証明するために開発したニューラルモデルを基に、Johnson らが計算機上でのシミュレーションが可能となるように拡張を施したものである[1][2]。PCNN は二次元空間上の画像を容易に扱うことが可能であり、また FFS も容易に導入できるという特徴を持つ。そこで、我々は PCNN に基づいた新たな SONN モデルを提案し、差分法による移動

物体抽出結果に含まれるノイズを除去し、移動物体の領域全体を抽出する方法を示す。本手法では、ピクセルの時空間における相関関係からノイズと移動物体を区別し、過渡応答性により移動物体の欠損個所の補完を行う。そのため空間的な特徴が同じであっても移動物体とノイズを判別することができる。

Feed Forward Shunting (FFS)

FFS は SONN の一種であるが、その構造を図 1(a)に、動作式を式(1)に示す。

$$X_i(t+1) = AI_i(t+1) + BI_i(t+1)X_{i-1}(t) \quad (1)$$

ここで、 $AI_i(t+1)$ は時刻(t+1)における入力であり、 X_{i-1} は左隣にあるニューロンの出力である。また、A,B は固定された結合重みである。

図 1(b)のように一次元にネットワークを組んだ場合の動作は以下の通りである。もし、入力 $I_i(t+1)$ が十分大きければ、 $X_{i-1}(t)$ は増幅されて X_i へ伝播する。しかし、大きくない場合は減衰して伝播する。増幅が生じるのは、 $BI_i(t+1)$ が 1 以上になる場合だけである。つまり、 $X_{i-1}(t)$ と相関のある入力 $I_i(t+1)$ がある場合のみ活動電位の伝播が生じる。このため、図 1(b)の入力のようにかたまりとなって移動する外部刺激はニューロン間を伝播するが、ランダムに変動する外部刺激が伝播されることはない。

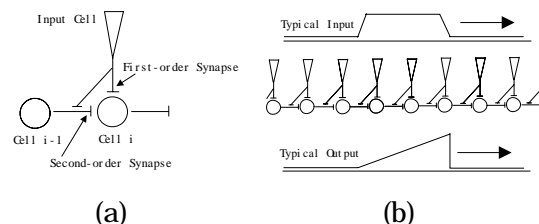


図 1 FFS の構造と一次元ネットワーク

(*) 慶應義塾大学 SFC 研究所 (Keio Research Institute at SFC)
(**) ニッタン株式会社 (Nittan Co.,LTD)
(***) 慶應義塾大学環境情報学部 (Faculty of Environmental Information Keio University)

Pulse Coupled Neural Network(PCNN)

ここでは二次元に配列されたニューロン(i,j)に着目して説明を行う。ニューロンは Dendritic Tree, Linking modulator, Pulse generator と呼ばれる三つの部分から構成される。Dendritic tree は, Feeding 部, Linking 部と呼ばれる二つの部分に分かれており, それぞれ Linking modulator に対して独立した出力を生成する。

Feeding 部と Linking 部は, シナプス結合 M, W を介して近傍のニューロンと接続されている。なお, 外部刺激 S は Feeding 部だけが受ける。また, Feeding 部と Linking 部はそれぞれ前の時刻の値を保持しており, その値は時間とともに減衰する。Feeding 部の値 F(t) と Linking 部の値 L(t) は次式よりそれぞれ求まる。

$$F_{ij}[t] = \exp(-\alpha_F \delta_t) F_{ij}[t-1] + S_{ij} + V_F \sum_{kl} M_{ijkl} Y_{kl}[t-1] \quad (2)$$

$$L_{ij}[t] = \exp(-\alpha_L \delta_t) L_{ij}[t-1] + V_L \sum_{kl} M_{ijkl} Y_{kl}[t-1] \quad (3)$$

ここで, S は外部刺激, $Y_{kl}[t-1]$ は (k,l) に位置する近傍ニューロンの前の時刻での出力を示す。F, L は時定数を, t は単位時間を表す。VF 及び VL はシナプスの重み行列を正規化する為の定数である。

Linking Modulator は, Dendritic tree の二つの部分の出力を結合強度 にしたがって結合して内部状態を決定する。ニューロンの内部状態は次式によって計算される。

$$U_{ij}[t] = F_{ij}[t] (1 + \beta L_{ij}[t]) \quad (4)$$

Pulse generator は, 内部状態と動的閾値を比較して出力を決定する。出力 Y は以下の式より導き出される。

$$Y_{ij}[t] = \text{step}[U_{ij}[t] - \theta_{ij}[t]] \quad (5)$$

ここで, step はステップ関数である。

また, θ_{ij} は動的閾値であり以下の様に定義される。

$$\theta_{ij}[t] = \exp(-\alpha_\theta \delta_t) \theta_{ij}[t-1] + V_\theta Y_{ij}[t] \quad (6)$$

ここで, V は正規化定数, θ_{ij} は時定数を表す。

ネットワークの 1 回の動作は, ニューロンを構成する三つの部分が下記の様に動作することで完了する。

1. それぞれの受容野の入力加重和を計算する(Dendritic tree)
2. Linking 部の出力により Feeding 部の出力を調整(Linking Modulator)
3. 出力パルスを生成(Pulse generator)
4. 発生したパルスが, 受容野を介して他のニューロンに入力される

また, PCNN を画像処理に適用するには, 画像中の 1 画素に対して PCNN の 1 ニューロンを割り当ててネットワークを動作させればよい。具体的には, 入力画素と同じ個数のニューロンを画素と同様に二次元に配置し, 正規化した画素の輝度値を対応するニューロンの外部刺激 S として入力する。なお, 出力としては, 正規化したそれぞれのニューロンの出力を画素の輝度値とみなせば, ネットワークの動作 1 回につき 1 枚の画像を得ることとなる。

移動物体抽出手法の提案

提案手法は, フレーム間差分法による輝度変化の抽出と PCNN を変更した SONN によるノイズ除去及び抽出領域の欠損補完からなる。フレーム間差分法を用いて動画から移動物体の全体を抽出するには, フレーム間差分法の出力に含まれるノイズを除去し, 抽出領域の欠損補完を行う後処理が必要となる。このような処理を行うために提案モデルでは PCNN の

Feeding 部, Linking 部, そして閾値に対して以下に述べるような変更を行った。

PCNN 基本モデルの Feeding 部は外部刺激と近傍ニューロンの出力を受けているが, 提案モデルでは外部刺激と近傍ニューロンの出力の相関を受けるように変更を行った。また, PCNN の基本モデルの外部刺激はニューロンが対応している画素の値となるが, 提案モデルでは対応する画素の近傍画素の値も考慮し, さらに外部刺激に対して過渡応答性を持つように変更を行った。式(2)を式(7), (8)に変更する。

$$F_{ij}[t] = P_{ij}[t] + \exp(-\alpha_F \delta_t) F_{ij}[t-1] + V_F \sum_{kl} M_{ijkl} Y_{kl}[t-1] S_{ij}[t] \quad (7)$$

$$P_{ij}[t] = V_p \sum_{kl} R_{ijkl} S_{ij}[t] + \exp(-\alpha_p \delta_t) P_{ij}[t-1] \quad (8)$$

ここで, Rijkl はシナプス結合重み, $\exp(-p t)$ は減衰項, VP は正規化定数を示す。

基本 PCNN において, Linking 部の値は Feeding 部の値を変調する役割を担っているが, 近傍ニューロンが発火していない場合に負の値を出力するように変更を行った。式(3)を式(9)と置き換える。

$$L_{ij}[t] = \exp(-\alpha_L \delta_t) L_{ij}[t-1] + V_L \sum_{kl} M_{ijkl} (Y_{kl}[t-1] - 1) \quad (9)$$

PCNN 基本モデルの動的閾値はニューロンに不応期を与えているが, 本研究の目的においては常に移動物体を抽出する必要があるので, 閾値は定数とした。式(6)を式(10)に変更する。なお, この変更により提案モデルは PCNN をベースとしてはいるが性質が異なったものとなる。そのため, 提案モデルは広義の意味において SONN モデルの一種と位置付けられる。

$$\theta_{ij}[t] = \text{const.} \quad (10)$$

提案モデルを用いた移動物体抽出手順を示す。

1. カラー動画画像から連続フレーム画像を生成
2. フレーム画像の輝度値を 256 階調で計算し, 輝度画像を生成
3. 生成した輝度画像を 0 から 1 に正規化
4. 正規化した輝度画像をフレーム間差分法で処理し, 輝度変化の画像を生成
5. 輝度変化の画像を SONN の外部刺激として入力しネットワークを動作
6. SONN の各ニューロンの出力 (1or0) と対応するカラーフレーム画像の画素の積を計算
7. 2.~6.の処理をフレーム毎に繰り返す

人工動画画像を用いた実験と考察

提案手法を実動画像に適用する前に, 人工動画画像を用いてモデルパラメータ値が結果に与える影響を確認する。実験は Pentium III processor 650 MHz を搭載した DOS-V 互換機上で, MATLAB6.5 を用いて行った。人工動画画像は, ランダムノイズを付加した黒い背景画像に, 屋外で車を撮影した画像から車の部分のみ手動で切り出した領域を, 1 ピクセルずつ位置を右にずらしながら重ね合わせて作ったフレーム画像からなる。ノイズの輝度値は一樣乱数により決定し, また, その位置は発生率に応じてフレーム毎にランダムに生成した。ここで扱うノイズは, 発生率と最大輝度によって決まるため, 以降, NOISE (発生率 (%), 最大輝度 (0~255)) と表現する。

入力データは輝度値に変化があるハッチバック車と NOISE(5%, max50)を付加した背景からなる連続画像 30 フレームである。使用したパラメータセットを表 1 にまとめる。なお, シナプス結合重み M, W, R には要素の総和を 1 に正規化したガウシアンカーネルを用いた。そこで, 表中ではガウシアンカーネルの半径 (rM, rW, rR), つまりいくつ迄離れたニューロンと結合しているかを示した。

表 1 パラメータセット一覧

パラメータ セット名	β	r_M	r_W	r_R	$\log 2/\alpha_L$	$\log 2/\alpha_F$	$\log 2/\alpha_P$	V_L	V_F	V_P	θ
PSet1	0	0	0	0	0.001	0.001	0.001	0	0	5	0.9
PSet2	0	1	0	0	0.001	0.001	0.001	0	5	5	0.9
PSet3	0	0	0	1	0.001	0.001	0.001	0	0	5	0.9
PSet4	0.25	1	3	1	0.4	1.2	0.3	4	100	50	0.9
PSet5	0.25	1	3	1	0.4	1.2	1.2	4	100	50	0.9
PSet6	0.25	1	3	1	0.4	2	2	4	100	50	0.9
PSet7	0.40	1	3	2	0.2	0.2	0.1	1.9	5	15	0.6

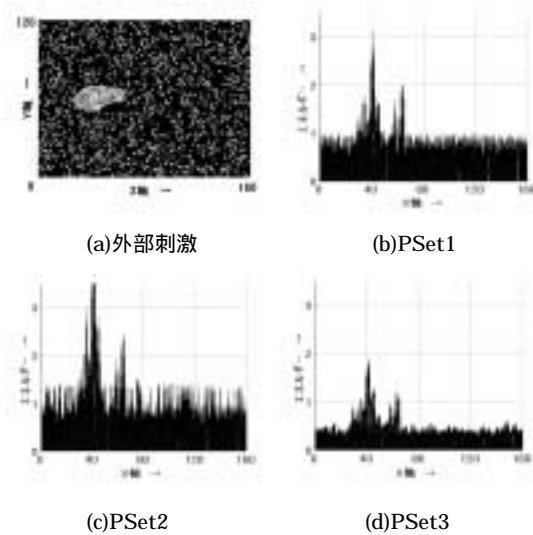


図 2 各ニューロンの Feeding 部の値 (30 フレーム目)



図 3 過渡応答特性が出力に与える影響

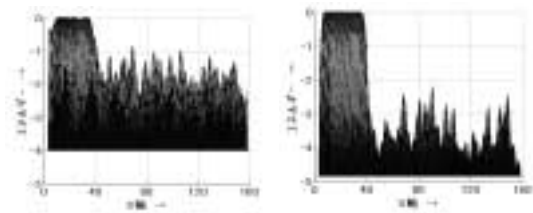


図 4 各ニューロンの Linking 部の値

各パラメータセットは、次の 5 つの点を確認する目的で設定している。なお、以下の実験結果は全て図 2(a)に示す 30 フレーム目の外部刺激を入力した際のものである。

1. 外部刺激の様子

パラメータ値 PSet1 に設定して動作させた際の各ニューロンの Feeding 部のエネルギー値 $F(t)$ を図 2(b) 示す。なお、結果は画像水平軸 x 画像垂直軸 y エネルギー軸の三次元で表したグラフを画像水平軸 x エネルギー軸平面に射影した二次元グラフで示した (以下同じ)。Feeding 部の式(7)(8)に PSet1 の値を代入すると $F_{ij}[t]=5 \cdot S_{ij}[t]$ となるから、与えられたノイズによる外部刺激の最大値は 50 であるため、Feeding の値に与えるノイズの影響は最大で $50/255 \times 5$ はおおよそ 1 と

なるはずである。図 2(b)において、ノイズ位置に当たるニューロンのエネルギー値は確かに最大 1 程度である。一方、移動物体の画素から外部刺激を受けているニューロンの値は最大 3 程度となっている。

2. 近傍からの出力を入力に関連付けたことによる効果

PSet2 に設定した際の $F(t)$ を図 2(c) に示す。Feeding 部の式に PSet2 の値を代入すると

$$F_{ij}[t] = 5 \cdot S_{ij}[t] + 5 \cdot \sum_{kl} M_{ijkl} Y_{kl}[t-1] S_{ij}[t]$$

ここで $S_{ij}/255 \times 5 > 1$ より、外部刺激として $S_{ij} > 50$ が与えられたニューロンが、出力が増幅されて伝播されるニューロンとなる。しかし、与えられたノイズによる S_{ij} の最大値は 50 であるので、近傍ニューロンの出力値 1 は全て 1 以下に抑制されて伝播される。また、空間的にランダムに出現するノイズが前の時間にあった場所の隣に出現する率は低いのでさらに抑制されて伝播されるはずである。図 2(c)より、Feeding 部に現れるノイズの最大値は 1.5 程度になっており、図 2(b)におけるノイズの最大値 1.0 からの増加分、つまり出力結合に入力に関連付けたことによる増分は 0.5 であるから、確かに近傍ニューロンの出力値 1 は減衰して伝播されている。つまり、ノイズの影響が抑制されていることが確認できる。一方、移動物体から外部刺激を受けているニューロンの F_{ij} の最大値はグラフには示されていないが 4.5 程度であり、出力結合に入力に関連付けたことによる増分は 1.5 となるから、近傍ニューロンの出力値 1 が増幅されて伝播されている。つまり、移動物体の影響は促進されているといえる。

3. 近傍からの外部刺激による効果

PSet3 に設定した際の $F(t)$ を図 2(d) に示す。図 2(d)を PSet1 の実験結果である図 2(b)と比較すると、Feeding の値が全体的に小さくなっているが、特にノイズ部分の減少が著しい。したがって、ノイズの影響を低減できたといえる。

4. 過渡応答特性の影響

PSet4, PSet5, PSet6 に設定して処理した出力結果をそれぞれ図 3 に示す。移動物体が適切に抽出されるように調整された時定数パラメータ PSet4 による処理結果である図 3(a)と時定数を少し大きくした PSet5 による処理結果である図 3(b)を比べると、図 3(b)に進行方向と逆側に尾ひれのようなものが抽出されている。さらに時定数を大きくした PSet6 による処理結果にはその尾ひれがより大きくなる現象がみられた(図 3(c))。この結果より、時定数を適切に設定すれば導入した過渡応答特性により抽出領域の欠損を補完する効果が期待できる。

5. Linking 部に行った改良の効果

PSet4 を設定し、3 フレーム目と 5 フレーム目を処理した際の各ニューロンの Linking 部の値 $L(t)$ を図 4 に示す。図 4 より Linking 部の値は、自分自身および近傍が前の時刻に発火しているところだけ 0 となっており、それ以外は負の値を出力することが確認できる。つまり、発火している領域の周辺の部分は、発火していないニューロンの影響により負の値を持つことになる。Linking 部の値が負である場合、式の変調作用によりニューロンの発火を抑える方向に働く。この作用によりノイ

ズを減少させる効果が期待できる。なお、図 4(a) (b)と処理が進むとノイズの影響で発火していたニューロンの Linking 部のエネルギーが負の方向に大きくなっていることが確認できる。

以上のことから基本 PCNN に対して行った変更によるそれぞれの効果の相互作用により、提案モデルは外部刺激からノイズを除去し、まとめて移動する外部刺激の欠損の補完を行う機能を有することが期待できる。

実動画像を用いたシミュレーション

実動画像データとして、Dataset for 2nd IEEE International Workshop on Performance Evaluation of Tracking and Surveillance(英国 Reading 大学作成、ftp://pets2001.cs.rdg.ac.uk/PETS2001/)を用いたシミュレーションを行った。入力としては、データセット名:dataset1,分類:test,カメラ:camera1のJPEGデータを365×274のサイズに変換したもののから、以下の2シーンを抜粋して用いた。SCENE1(フレーム番号:450-600[5フレーム毎に入力])(図5)。SCENE2(フレーム番号:2490-2640[5フレーム毎に入力])(図6)。シミュレーションの方法としては、提案手法に関しては、パラメータを表1のPSet7に設定し処理を行った。また、従来手法としては、フレーム間差分(閾値=0.1)処理後、次示すモルフォジフィルタ処理を行った[4]。

$$S_k = [(R_k < c > C_3) < o > C_3] < c > C_5$$

ここで、 R_k は原画像、 S_k は処理結果、 C_3, C_5 は直径がそれぞれ3,5の円形オペレータを表す。



図 5 SCENE1 の例 (左から 1,15,30 フレーム)



図 6 SCENE2 の例 (左から 1,15,30 フレーム)

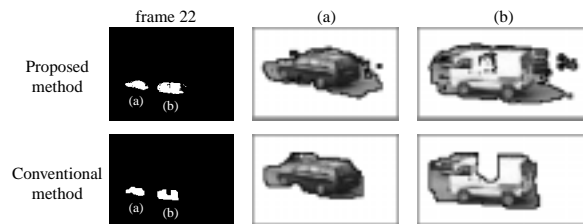


図 7 SCENE1 の結果 (上:提案手法, 下:従来手法)

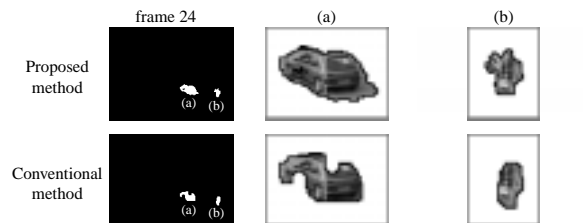


図 8 SCENE2 の結果 (上:提案手法, 下:従来手法)

SCENE1の22フレーム目とSCENE2の24フレーム目の結果をそれぞれ図7,8に示す。まず、SCENE1の結果である図7より、両手法とも大きなノイズは抽出されていない。ところが、従来手法による抽出物体(b)の車体が大きく欠損してしまっているが、提案手法の出力ではその欠損がほとんど見られないことが分かる。また、本論文の実験結果としては示していないが、従来手法の結果には欠損しているフレームと欠損なく抽出されているフレームが混在していた。これは、移動中に車両の向きが変わっているため光のあたり具合が変化した影響を受けたものと考えられる。一方提案手法の結果は安定して移動物体を抽出していた。次にSCENE2の結果である図8より、ノイズに関してはSCENE1と同じである。そして従来手法の抽出物体(a)の車体には欠損している部分があるが、提案手法の出力には、ほとんど欠損が見られないことが分かる。以上の結果より、提案手法は従来手法に比べノイズ除去能力に優れ、さらに抽出物体の欠損を補完する能力にも優れていることが理解できる。これは、従来手法がピクセルの空間的な関係に基づいてノイズ除去と欠損の補完をおこなっているため、対象の輝度変化に処理結果が影響されやすいのに対し、提案手法はピクセルの時空間的な関係に基づいて処理を行うため、対象の輝度変化に影響されにくいと考えられる。なお、これらのシミュレーションに要したフレーム毎の実行時間は、提案手法が約4秒、従来手法で約6秒であった。したがって、提案手法の方が実行速度の面で30%短縮できた。

本手法には、以下の点について問題がある。(1)処理の初期段階では、出力が安定しない。(2)移動物体の後ろに尾ひれのようなものが抽出されやすい。(3)対象物の画像における見かけ上の移動速度に対して、十分短い時間間隔で処理されない、抽出物の輪郭が不明瞭になる。

おわりに

本論文では、フレーム間差分とPCNNを基にしたSONNを用いた移動物体領域の抽出手法を提案した。提案手法は動画像からノイズを除去し移動物体の全体を抽出できることが、人工画像による実験と実画像を用いたシミュレーションにより確認できた。また、本研究で行ったシミュレーションでは、実行時間が従来手法より約30%短縮されることが確認された。今後は、動画像認識システムの前処理に本手法を適用し、システム全体に与える提案手法の有効性を検証していきたい。

参考文献

- [1] Luis R.Lopez, Feedforward shunting: a simple second-order neural network motion sensor, Proc. SPIE, Vol.1297, p.350-358(1990)
- [2] R.Eckhorn, H.J.Reitboeck, M.Arndt, and P.Dicke, 1990, Feature linking via synchronization among distributed assemblies: simulations of results from cat cortex, Neural Comput. 2, 293-307.
- [3] John L. Johnson, 1994, Pulse-coupled neural nets: translation, rotation, scale, distortion, and intensity signal invariance for images, APPLIED OPTICS, Cvol. 33, No. 26, 6239-6253
- [4] J.Heikkila and O.Silven, A Real-Time Tracker for Visual Surveillance Applications, Proceedings 1st Int. Workshop on PETS, p.8-13(2000)