

時空間輪郭積分法による動柔物体の抽出

鈴木 智[†] 中根 一成[†]
ジャック・スクランスキー^{††}

本論文では、動画像から動柔物体を抽出する新しい方法を述べる。一つの動物体は、なめらかに濃度変化をしている領域（パッチと呼ぶ）に分割できるので、まず、この領域の動きによって生じる変化領域（動領域と呼ぶ）を時間エッジの位置と濃度値から抽出する。次に、動領域の重心の時間的変位から動領域の動きを推定する。この動領域解析は、重心の位置を変えない変形に対して不変なオプティカルフローを推定できる。次に、動領域の境界の時間的な変化からパッチの部分輪郭を動輪郭として抽出する。動輪郭は、動物体の位置の手がかりを与える。最後に、動輪郭をその動き特徴によりクラスタリングすることにより、動物体を抽出する。本方法を、複雑な背景上に多数の歩行者が不均一な動きをしている横断歩道の動画像に適用した結果、約 86% の歩行者を抽出できた。提案した方法は、背景を抽出する必要がなく、カメラを固定しておく時間が短時間（上記の実験では 1.23 秒）でよいという長所がある。

Extracting Nonrigid Moving Objects by Spatio-temporal Contour Integration

SATOSHI SUZUKI,[†] KAZUNARI NAKANE[†] and JACK SKLANSKY^{††}

We describe a new method for extracting nonrigid moving objects. Every moving object may be approximated by a set of smooth-intensity patches. First, our method finds variant regions which are caused by the motions of these patches. The variant regions are extracted by integrating the locations and the intensities of temporal edges. Next, optical flows are estimated as temporal displacements of gravity centers of variant regions. This estimation is invariant for the deformation that does not change the locations of the gravity centers. Then, motion edges are extracted from the subboundary of each variant region, which gives us the location of a moving object. Finally, the motion edges are clustered spatially based on their motion to extract objects. The performance of the proposed method was tested successfully on extracting pedestrians from a cinematic sequence viewing a street crossing. The extracted rate was 86%. The proposed method needs no background extraction. Furthermore, our method requires a camera to be stationary only for 1.23 sec.

1. ま え が き

動画像を用いて動物体を抽出する方法について述べる。動物体の抽出・追跡は、監視システム、自動車等の移動物体のナビゲーション、医用画像解析などに有効である^{1)~5)}。動物体の抽出方法として、多くの方法が提案されている。これら进行分类すると、背景抽出を行い、その差分として動物体を抽出する方法^{2)~4)}と、背景抽出を行わず、動き情報から直接動物体を抽出する方法^{5)~7)}に大別できる。前者の方法は、背景抽出の

ためにカメラを長い時間静止しておかなければならない。また、天候の変化による背景の濃度値変化に弱い。一方、後者の方法では、動き情報の抽出のためにオプティカルフロー推定法が多数提案されているが^{8)~12)}、雑音や環境変化等により、一つの動物体に属する画素が不均一なオプティカルフロー推定値を持つ。そのため、個々の動物体を抽出することは難しいという問題がある。特に、人間のように非剛体で動く物体（動柔物体と呼ぶ）の抽出は、非剛体性によりオプティカルフローが不均一になるため困難である。

本論文は、背景抽出を行わず、動き情報から動物体を抽出する一方法を述べる。まず、従来方法の問題点を解決するため、オプティカルフローの推定法として時空間輪郭積分を用いる新しい方法（時空間輪郭積分

[†] NTT ヒューマンインタフェース研究所
NTT Human Interface Labs.

^{††} カリフォルニア大学アーバイン校
University of California, Irvine

法と呼ぶ)を提案する。この方法で得られたオプティカルフローは、大局的で、かつ均一であることを実験的に示す。

時空間輪郭積分法では、まず、なめらかに濃度変化をしている領域(パッチと呼ぶ)の動きによって生じる変化領域(動領域と呼ぶ)を時間エッジ⁸⁾の位置と濃度値から抽出する。次に、動領域の重心の時間的変位から動領域のオプティカルフローを推定する。この動領域解析は、重心の位置を変えない変形に対して不変なオプティカルフローを推定できる。

時空間輪郭積分法と同様に、時空間から動き特徴を抽出する方法を山本がすでに提案している¹³⁾。しかし、剛体の等速水平運動という強い仮定のもとで、時空間の断面において空間的エッジの流跡を線分として求め、あらかじめ与えられた動物体領域を追跡したものであり、動物体の自動抽出には至っていない。

次に、動物体の輪郭から、動きに対して平行でない輪郭部分(動輪郭と呼ぶ)を抽出する方法を述べる。動輪郭は、時空間輪郭積分法により抽出された動領域とオプティカルフローから抽出される。動輪郭は、動物体の位置の手がかりを与える。そこで、動輪郭をその動き特徴によりクラスタリングすることにより、動物体を抽出する。

最後に、複雑な背景上を多数の歩行者が不均一な動きをしている横断歩道の動画をを用いた実験結果を示す。カメラを1.23秒だけ静止して置き、歩行者の抽出率約86%という結果を得た。

2. 時空間輪郭積分法によるオプティカルフロー推定

2.1 原理

まず、時空間輪郭積分法の概要を述べる。図1(a)に示すように、入力データは、動画であり、2次元画像の列である。動物体の動きにより、図1(b)において黒く塗りつぶした領域のように濃度値が変化する領域が存在する。この領域を動領域と呼ぶ。図1では簡単のため、典型的な一つの動領域だけを示している。この領域は、動物体の動きに平行でない輪郭の位置を時間で積分したものと見なすことができる。したがって、図1(c)のように、各フレームにおける輪郭の位置を描くと、等時間曲線を描くことができる。図1(d)のように、この等時間曲線で、動領域を2分割して、それらの重心の変位を求め、得られた変位を時間で正規化することによりオプティカルフローを得

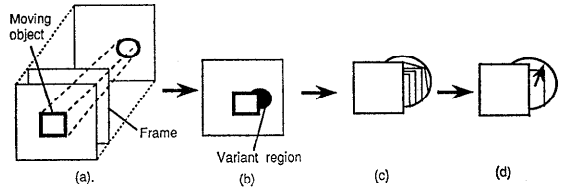


図1 時空間輪郭積分法の概要

(a) 動画像。(b) 動領域。(c) 動物体の輪郭の時間的位置。(d) 2分割された動領域の重心の変位。

Fig. 1 Spatio-temporal contour integration method for optical flow estimation.

(a) Motion image. (b) Variant region. (c) Temporal locations of moving object contour. (d) Displacement of gravity centers of divided variant regions.

る。

この時空間輪郭積分法の妥当性は、数式により導くこともできる。オプティカルフローベクトル (u, v) をもつ動物体のパッチの部分領域がフレーム #1 において二つの画素 (x_1^1, y_1^1) と (x_2^1, y_2^1) からなるとする。ここで、上つき添え字は、フレーム番号を表す。短い時間間隔 dt に対して動物体領域内でのオプティカルフローは一定である仮定すると、次式(1)が成り立つ。

$$\frac{dx}{dt} = \frac{x_1^{1+dt} - x_1^1}{dt} = \frac{x_2^{1+dt} - x_2^1}{dt} \quad (1)$$

ここで、 $dx = udt$ である。

この式が表す二つの式を加えて、展開すると次式(2)を得る。

$$\begin{aligned} \frac{dx}{dt} &= \frac{1}{2} \sum_{k=1}^2 \frac{x_k^{1+dt} - x_k^1}{dt} \\ &= \frac{1}{dt} \left(\frac{1}{2} \sum_{k=1}^2 x_k^{1+dt} - \frac{1}{2} \sum_{k=1}^2 x_k^1 \right) \end{aligned} \quad (2)$$

パッチの動き (u, v) は、フレーム1からフレーム $N+dt$ の間、不変であると仮定し、式(2)のフレーム対 $(1, 1+dt)$ をフレーム対 $(t, t+dt)$ 、 $t=1, 2, \dots, N$ に置き換えて得られる式の各両辺を加えていくと、次の式(3)が導出できる。

$$\frac{dx}{dt} = \frac{1}{dt} \left(\frac{1}{2N} \sum_{k=1}^2 \sum_{t=1}^N x_k^{t+dt} - \frac{1}{2N} \sum_{k=1}^2 \sum_{t=1}^N x_k^t \right) \quad (3)$$

この式は、一般性を失うことなく、 M 個の画素からなるパッチの部分領域に関する式(4)に拡張できる。

$$\frac{dx}{dt} = \frac{1}{dt} \left(\frac{1}{MN} \sum_{k=1}^M \sum_{t=1}^N x_k^{t+dt} - \frac{1}{MN} \sum_{k=1}^M \sum_{t=1}^N x_k^t \right) \quad (4)$$

dt の値を N にすることができると仮定すると、次式

が得られる.

$$\frac{dx}{dt} = \frac{1}{dt} \left(\frac{1}{MN} \sum_{k=1}^M \sum_{t=N+1}^{2N} x_k^t - \frac{1}{MN} \sum_{k=1}^M \sum_{t=1}^N x_k^t \right) \quad (5)$$

同様にして $dy = vdt$ とすると dy/dt も式(6)のように表現できる.

$$\frac{dy}{dt} = \frac{1}{dt} \left(\frac{1}{MN} \sum_{k=1}^M \sum_{t=N+1}^{2N} y_k^t - \frac{1}{MN} \sum_{k=1}^M \sum_{t=1}^N y_k^t \right) \quad (6)$$

式(5)と(6)から、パッチの動きは、フレーム1からフレーム N の間に变化した領域の重心と、フレーム $N+1$ からフレーム $2N$ の間に变化した領域の重心との変位から推定できることを示している. すなわち、時空間輪郭積分法により、オプティカルフローが推定できる. こうして、オプティカルフローの推定は、動領域の抽出問題に帰着する.

ところで、式(5)と(6)を用いたオプティカルフロー推定は、回転や重心の位置を変えない動物体の変形に対して不変である. したがって、本方法は、このような変形特性をもつ動柔物体に対して、局所的な変形は検出できないが、大局的なオプティカルフローが得られるという特徴がある.

2.2 時間エッジ検出

動領域の抽出のため、時間エッジ⁸⁾を用いる. 局所的な濃度値の差分から動領域を抽出する方法が考えられるが、变化したか否かを決定する閾値の設定が難しい. そこで、時間エッジを用いることにした.

時間エッジは、動画の各画素 (x, y) の濃度値の時間変化 $g(x, y, t)$ にガウス関数の2階微分をたたみこみ、時間軸方向の零交差点を時間エッジとして抽出する⁸⁾. 例えば、画素 (x, y) の濃度値 $g(x, y, t)$ が図2に示すように時間変化する場合、フレーム $t_1, t_2, t_3, t_4, t_5, t_6, t_7$ において時間エッジは検出される.

しかし、この方法を実データに適用すると雑音のために、偽の時間エッジを抽出してしまう. そこで、次の二つの処理を追加した. 一つは、 $g(x, y, t)$ のかわりに 3×3 近傍における濃度値の平均値を用いて雑音の

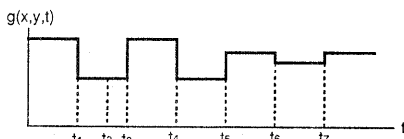


図2 時間エッジ ($t_1, t_2, t_3, t_4, t_5, t_6, t_7$) において検出される.
Fig. 2 Temporal edges occurred at frames $t_1, t_2, t_3, t_4, t_5, t_6$ and t_7 .

高周波成分を抑制する. もう一つは、いったん、得られた時間エッジに基づいて、隣接する時間エッジの間の区間について濃度値の平均値を求める. そして、隣接する区間の濃度平均値の差分が閾値 ϵ_1 以下ならば、これらの区間には含まれている時間エッジを偽の時間エッジとする. 例えば、図2では、フレーム t_6, t_7 における時間エッジは偽の時間エッジとなる. この閾値処理は大局的であり、局所的な濃度値の差分の閾値処理より閾値を容易に決められる. 図3は、実データ (128 画素 \times 120 画素 \times 41 フレーム) から時間エッジを抽出した結果を時空間の $x-t$ 断面 ($y=49$) について表示している. 動物体は斜めの傾きを持った領域として $x-t$ 断面に現れる. この領域の輪郭が、上記の方法で時間エッジとしてうまく抽出できることがわかる.

2.3 動領域解析

時間エッジ検出により、画像全体の動領域の和集合は時間エッジの位置の時間変化として得られた. しかし、まだ、各パッチの動きによりできる個々の動領域に分割されていないので、パッチのオプティカルフローを求めることはできない. 本節では、個々の動領域の抽出方法を述べる.

フレーム t_2 におけるオプティカルフロー推定を行うための動領域を抽出する. まず、次の二つの関数 $a(x, y, t_2)$ と $b(x, y, t_2)$ を計算する. $a(x, y, t_2)$ は、各画素 (x, y) において、フレーム t_2 の後で最初に時間エッジが生じるフレームまでの時間 (すなわち、フレーム数) である. 例えば、図2に示す画素では $a(x, y, t_2)$ は $t_3 - t_2$ となる. ただし、観測の時間 W を定め、その時間内に時間エッジが検出されない場合、時間エッジは見つからなかったとする. 動きによる変化が生じていない画素では、このように時間エッジは検出されない.

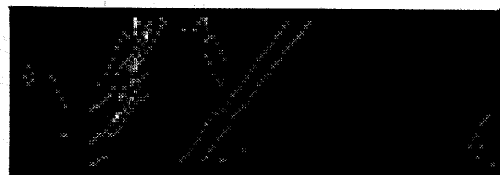


図3 実時空間画像の $x-t$ 断面に対する抽出された時間エッジ (\times で表す)
水平値が x 軸、垂直軸が t 軸である.

Fig. 3 Extracted temporal edges (denoted by crosses) for $x-t$ image.
The x -axis is horizontal and the t -axis is vertical.

$b(x, y, t_2)$ は、フレーム t_2 の後で最初に時間エッジが生じるフレームと 2 番目に時間エッジが生じるフレームとの間の画素 (x, y) の濃度値の平均である。

図 4 (a) に示すパッチの動きに対して、 $a(x, y, 1)$ と $b(x, y, 1)$ を計算すると、それぞれ図 4 (b), (c) のようになる。

動領域は、関数 $a(x, y, t_2)$ が増加し、かつ $b(x, y, t_2)$ がなめらかに変化している画素 (x, y) の領域として抽出できる。まず、 $a(x, y, t_2)=1$ である画素の連結領域を求め、さらに $b(x, y, t_2)$ の値がなめらかである連結領域に細分する。ただし、隣接する画素との差分が閾値 ϵ_2 以下であるならば、なめらかであると判定する。

次に、 $a(x, y, t_2)=2$ である画素の連結領域を求め、さらに $b(x, y, t_2)$ の値がなめらかである連結領域に細分する。そして、 $a(x, y, t_2)=1$ の連結領域に、 $a(x, y, t_2)=2$ の連結領域を統合する。 $a(x, y, t_2)=2$ である連結領域 R を、 $a(x, y, t_2)=1$ のどの領域に統合するかは次のように決める。 R に属する画素 (i, j) の近傍画素 (p, q) が $a(x, y, t_2)=1$ である連結領域に属し、かつ (i, j) と (p, q) が $b(x, y, t_2)$ についてなめらかである画素対 (i, j) と (p, q) をすべて求める。そして、最大の $a(p, q, t_2)$ を与える領域に、 R を統合する。これ

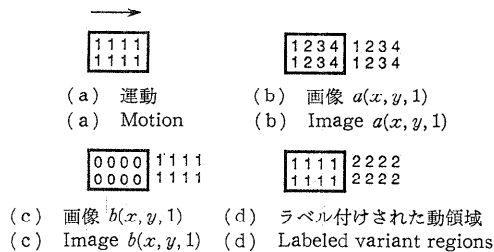


図 4 フレーム 1 における動領域の抽出 (パラメータ $W=4$ に設定した場合)

左から右へ 1 画素/フレームの速度でパッチ (濃度値は 1) は移動する。背景の濃度値は 0 である。長方形はパッチのフレーム 1 における位置を示す。(a), (d) では、空白は値 0 の背景を表す。(b), (c) では、空白は関数の値が不定である画素を表す。

Fig. 4 Variant region extraction at frame #1 obtained by setting W as 4.

The rectangles show the location of the moving patch at frame #1. The patch having intensity one moves from left to right at a speed of 1 pixel/frame. The background has intensity zero. In Figs. (a) and (d), blanks denote background pixels. In Figs. (b) and (c), blanks denote pixels whose function values are undefined.

は、時間的に最もなめらかな領域に統合することを意味する。

次に、 $a(x, y, t_2)=3$ である画素の連結領域を求め、さらに $b(x, y, t_2)$ の値がなめらかである連結領域に細分する。そして、 $a(x, y, t_2) \leq 2$ の連結領域に、 $a(x, y, t_2)=3$ の連結領域を統合する。以下、同様にして、領域を成長させる。

図 4 (d) は、図 4 (b) と (c) に示した $a(x, y, 1)$ と $b(x, y, 1)$ を用いた、動領域の抽出結果である。

各動領域に対して、式 (5) と (6) を適用して、フローを推定する。まず、動領域の $a(x, y, t_2)$ の最小値 T_1 と最大値 T_2 を求める。式 (5) と (6) における N を $(T_1+T_2)/2$ と設定する。そして、 $a(x, y, t_2)$ が T_1 から N までの領域と $N+1$ から T_2 までの領域に動領域を 2 分して、それぞれの重心の変位を時間 $(T_2-T_1+1)/2$ で正規化することにより動領域のオプティカルフローを得る。

2.4 問題点とその対策

前節で述べた方法には以下の四つの問題がある。

第 1 は、図 5 (a) に示すように、等しい濃度値でかつ等しい速度を持つ二つの動物体が正面衝突した場合に、二つの動領域が統合されて、一つの動領域として抽出されてしまうという問題である。

第 2 は、図 5 (b) に示すような場合、一つの動領域の一部が他の動領域や画像の縁により隠蔽されてしまうという問題である。

この二つの問題は、オプティカルフローは時間的になめらかに変化するという拘束をいれることにより抑制できる。すなわち、前節で述べた方法で動領域が得



図 5 動領域抽出に問題が生ずる場合。(矢印は移動軌道を表す)

(a) 濃度値が等しく、かつ速度が等しい動物体が正面衝突すると、二つの動領域が一つの動領域として抽出される。(b) 移動軌道が交差すると、左側の動物体の動領域がひずむ。

Fig. 5 Critical cases for extracting variant regions (arrows denote motion trajectories of objects). (a) Over-merging of variant regions occurs at the collision point of the trajectories. (b) Variant region of lefthand object is distorted at the collision point of the trajectories.

られたのち、オプティカルフローを以下のように求める。まず、動領域の $a(x, y, t_2)$ の最小値 T_3 を求める。 $a(x, y, t_2) \leq T_3 + 1$ の画素を連結領域に分割する。次に、各連結領域について $a(x, y, t_2)$ の最小値 T_1 と最大値 T_2 を求める。 N を $(T_1 + T_2)/2$ と設定して、式(5)と(6)を適用して各連結領域のオプティカルフローを得る。次に、 $a(x, y, t_2) \leq T_3 + 2$ の画素を連結領域に分割する。同様に、各連結領域のオプティカルフローを計算する。一つ前の反復で得られた連結成分と今回得られた連結成分で包含関係にあるものについて、オプティカルフローの変化がなめらかでなくなった連結領域を求める。その連結領域の総面積がすべての連結領域の総面積の $\varepsilon_3\%$ 以上になったとき、処理を終了する。ただし、オプティカルフローの角度の差分が閾値 ε_4 以下であるときなめらかであると判定する。処理が終了しない場合には、次は $a(x, y, t_2) \leq T_3 + 3$ を満足する画素の連結成分について同様の処理を行う。以下同様に処理を進め、処理が終了したとき、一つ前の反復で得られたオプティカルフローを最終的フローとする。

第3の問題は、 $a(x, y, t_2)$ の最小値と最大値が等しい動領域に対して、式(5)と(6)が適用できないということである。この問題は図6に示す小さなパッチからなる動物体について、図6(d)の動領域2および3に生じる。この問題は前節で述べた動領域抽出の後、この問題が生じている動領域を一つずつ、次のように再成長させることにより解決できる。まず、次の二つの関数 $c(x, y, t_2)$ と $d(x, y, t_2)$ を計算する。 $c(x, y, t_2)$ は、各画素 (x, y) において、フレーム t_2 の後で2番目に時間エッジが生じるフレームまでのフレーム数である。 $d(x, y, t_2)$ は、フレーム t_2 の後で2番目に時間エッジが生じるフレームとの間の画素 (x, y) の濃度値の平均である。図6(a)に示すパッチの動きに対して、 $c(x, y, 1)$ と $d(x, y, 1)$ を計算すると、それぞれ図6(e), (f)のようになる。最初の領域成長の結果を初期値として、関数 $a(x, y, 1)$ と $b(x, y, 1)$ のかわりにそれぞれ関数 $c(x, y, 1)$ と $d(x, y, 1)$ を用いて図6(g)と(h)に示すように再成長させる。再成長は、式(5)と(6)が適用可能になったところで終了する。この再成長を成長がおきなくなるまで続けると、再成長させた領域と再成長させなかった領域との統合が必要になる。そこで、処理を簡単にするため、式(5)と(6)が適用可能になった時点で再成長を終了する。再成長された動

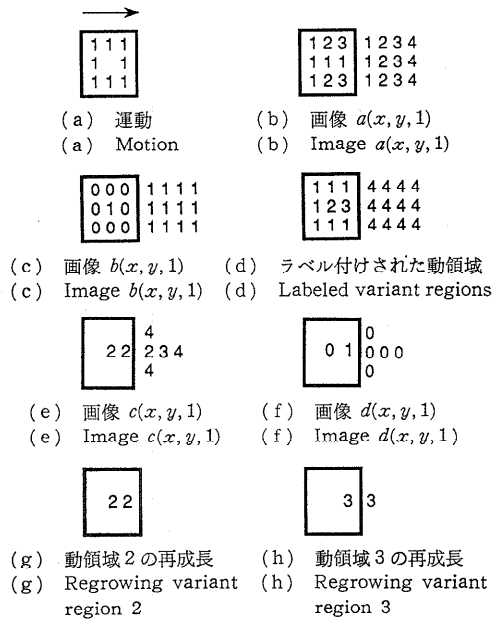


図6 フレーム1における動領域の抽出 (パラメータ $W = 4$ に設定した場合)

左から右へ1画素/フレームの速度でパッチ (濃度値は1) は移動する。背景の濃度値は0である。長方形はパッチのフレーム1における位置を示す。(a), (d), (g), (h)では、空白は背景画素を表す。(b), (c), (e), (f)では、空白は関数の値が不定である画素を表す。

Fig. 6 Variant region extraction at frame #1 obtained by setting W as 4. The rectangles show the location of the moving patch at frame #1. The patch having intensity one moves from left to right at a speed of 1 pixel/frame. The background has intensity zero. In Figs. (a), (d), (g) and (h), blanks denote background pixels. In Fig. (b), (c), (e) [and (f)], blanks denote pixels whose function values are undefined.

領域のオプティカルフローの計算は、再成長により加えられた画素 (x, y) の $a(x, y, 1)$ として $c(x, y, 1)$ を用いることにより、前節で述べた方法で計算できる。

第4の問題は、本来一つの動領域が分割されてしまう場合があるということである。この問題は、撮影された画像上で、動物体のフレーム間の移動距離が動物体の動き方向の長さより長い場合と、背景の一部分が動物体に等しい濃度値を持っている場合とに生じる。この問題は、上記の処理で得られた動領域を統合することにより対処可能であるが、処理が複雑になる。そこで、カメラの撮影スピードは、第1の場合が生じない程度に高速であると仮定する。また、第2の場合

は、非常に稀な場合であるので、本論文では第2の場合が生じないと仮定する。

3. 動物体の抽出

3.1 動輪郭の抽出

動輪郭とは、動物体をなすパッチの輪郭で動物体の動きに平行でない部分である。動輪郭のおよその位置は、動領域の輪郭で、パッチの輪郭の初期位置に隣接する部分を求めることにより得られる。具体的には、各動領域 R とそのオプティカルフロー (u, v) から、以下のようにして動輪郭を抽出する。ベクトル (u, v) に方向が最も近いベクトルを $(-1, -1)$, $(-1, 0)$, $(-1, 1)$, $(0, -1)$, $(0, 1)$, $(1, -1)$, $(1, 0)$, $(1, 1)$ の中から選び、それを (u', v') と表すと、画素 $(x-u', y-v')$ が動領域 R 上にないならば、 (x, y) を動輪郭上の画素とする。したがって、動輪郭は画素の集合となる。

図7は、移動している一人の歩行者（黒い部分）から抽出された動輪郭を示している。等しい数字をもつ画素の集合が同一の動輪郭に属する。図7から動輪郭が動物体の位置を与えることがわかる。

3.2 動輪郭の空間的クラスタリング

動物体は、動輪郭の集合として表現できる。そこで、動輪郭を、それに対応する動領域の動き特徴を用いてクラスタリングすることにより、個々の動物体を抽出する。ただし、動物体に関する知識として、動物体の水平方向の長さの上限 L_h と垂直方向の長さの上限 L_v が与えられていると仮定する。

フレーム t_2 における動物体抽出方法を説明する。

一つの動物体は、パッチ（なめらかに濃度変化をしている領域）の集合とみなすことができる。各パッチに対して、その移動により、移動の前方に一つの動領域ができ、移動の後方に一つの動領域ができる⁷⁾。さらに、各動領域に対して一つの動輪郭が抽出される。したがって、動輪郭を簡易な方法でクラスタリングすると、同じ濃度値をもつ複数の動物体が接近している場合、それらが一つの動物体として抽出されてしまう。そこで、動物体の輪郭に位置する動輪郭（最外郭動輪郭と呼ぶ）をまず求める。

一つのパッチの移動の後方にできる動領域 V_1 とその動輪郭を E_1 、前方にできる動領域 V_2 とその動輪郭を E_2 と表すと、これらの動輪郭は次の条件1から5を満足する。すなわち、二つの動輪郭は同じオプティカルフローを持ち（条件1）、二つの動輪郭の距離はパッチの大きさにより拘束され（条件2と3）、二つの動輪郭に対応する動領域は対応するパッチの濃度値があらわれる時間があり（条件4）、かつ二つの動輪郭の位置関係はオプティカルフローにより決まる（条件5）。また、逆に条件1から5をすべて満足すれば、ほぼ、 E_1 と E_2 は一つのパッチの移動の前後にできる動輪郭の対であると考えられる。

条件1) E_1 のオプティカルフローと E_2 のオプティカルフローとの角度の差分が閾値 ε_5° 以下であり、かつ大きさの差分が閾値 ε_6 以下である。

条件2) E_1 を含むクラスタと E_2 を含むクラスタを統合したクラスタの水平長が L_h 以下かつ垂直長が L_v 以下である。

条件3) 動輪郭 E_1 の重心と動輪郭 E_2 の重心との距離が閾値 ε_7 以下である。

条件4) 関数 $e(x, y, t_2)$ の動領域 V_1 上の平均値と関数 $b(x, y, t_2)$ の動領域 V_2 上の平均値の差分が閾値 ε_8 以下である。ここで、 $e(x, y, t_2)$ は、フレーム t_2 の前後で最初に時間エッジが生じるフレーム間の画素 (x, y) の濃度値の平均である。

条件5) 動領域 V_1 の重心から動領域 V_2 の重心への方向と動輪郭 E_1 のオプティカルフローおよび E_2 のオプティカルフローの方向との差分が閾値 ε_9° 以下である。

動輪郭が最外郭動輪郭でなければ、その動輪郭をパッチ移動の後方にできる動輪郭とするパッチと、パッチ移動の前方にできる動輪郭とするパッチが両方

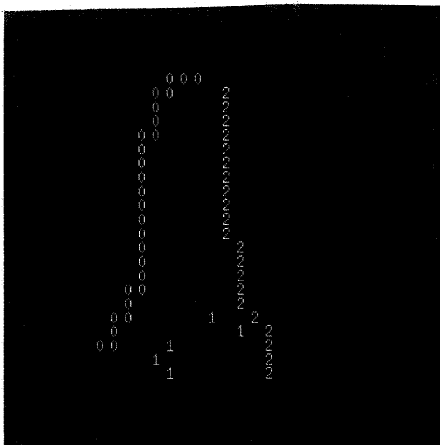


図7 ラベル付けられた動輪郭
各数字は画素が属する動輪郭のラベルである。
Fig. 7 Labeled motion edges.
Each digit shows a motion edge.

とも存在する。一方、最外郭動輪郭がパッチの移動の後方にできる動輪郭である場合、その動輪郭を移動の前方とするパッチは存在しない。その他の場合、すなわち、最外郭動輪郭がパッチの移動の前方にできる動輪郭である場合、その動輪郭を移動の後方とするパッチは存在しない。したがって、条件1から5を用いて、対をなす動輪郭が一方にしか存在しない動輪郭を抽出することにより、最外郭動輪郭を抽出する。

次に、各動輪郭を一つのクラスタに割り当てる。そして、すくなくともどちらか一方が最外郭動輪郭である動輪郭対 (E_1, E_2) の中で、条件1から5をすべて満足し、かつ式(7)を最小にする動輪郭対 (E_1, E_2) を求め、 E_1 を含むクラスタと E_2 を含むクラスタを一つのクラスタに統合する。この処理を統合が生じなくなるまで反復する。

$$S(E_1, E_2) = S_{12} - S_1 - S_2 \quad (7)$$

ここで、 S_1 は E_1 を含むクラスタの外接長方形の面積、 S_2 は E_2 を含むクラスタの外接長方形の面積、 S_{12} は S_1 と S_2 とを統合してできるクラスタの外接長方形の面積である。

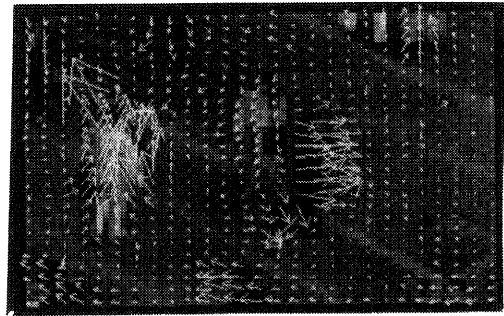
最後に、動輪郭対 (E_1, E_2) の中で、条件1から3のすべてを満足し、かつ式(7)を最小にする動輪郭対 (E_1, E_2) を求め、 E_1 を含むクラスタと E_2 を含むクラスタを一つのクラスタに統合する。この処理を統合が生じなくなるまで反復する。

このように、最外郭動輪郭を含む動輪郭対の統合を行った後、制約をゆるくした動輪郭対の統合を行うことにより、信頼性の高い統合から低い統合へと処理が進み、良いクラスタリング結果を得ることができる。

4. 実験

文献3) で用いられた横断歩道の動画像に、本方法を適用した結果を述べる。この動画像は CCD カメラを固定して撮影したもので、約 40 人の歩行者が不均一な動きをしているものである。隣接する歩行者間では、それらが重なり合いオクルージョンが生じている。8秒間の間に 240 フレームが撮影された。各フレームは 256×240 画素のサイズを持つ。本実験では、処理時間の短縮のためフレームのサイズを単純な 2×2 画素の平均で $128 \text{ 画素} \times 120 \text{ 画素}$ に削減したものをを用いた。

図8(a)は、従来の典型的な方法である文献9)の方法で、フレーム100とフレーム102からオプティカルフローを推定した結果である。ただし、この方法



(a) ホーンらの方法
(a) Horn & Schunck method



(b) 本方法
(b) Proposed method

図8 フレーム100において推定されたオプティカルフロー

Fig. 8 Estimated optical flows at frame #100.

のパラメータ $\alpha^2=10$ として反復回数は 100 回とした。図8(b)は、時空間輪郭積分法によりフレーム91からフレーム127までの37フレーム(1.23秒に対応する)を用いてフレーム100におけるオプティカルフローを推定した結果である。ただし、本方法のパラメータ $W=19$ 、ガウス関数の標準偏差 = 3 画素、閾値 $\varepsilon_1=5$ 、 $\varepsilon_2=10$ 、 $\varepsilon_3=40\%$ 、 $\varepsilon_4=90^\circ$ とした。この図から、時空間輪郭積分法で得られるフローは大局的で、かつ均一であることがわかる。また、動物体の境界が明確であり、静的領域のフローは0になるという特長がある。図8に示したものは、実験結果の一部であり、フレーム101から120までにおけるフロー推定でも同様の傾向を示した。提案した方法は、画素単位の精度のフローは、得られないが、動物体の抽出や追跡には十分であると考えられる。

文献9)の方法は、2枚のフレームだけを用いるのに対して、提案した方法は37枚という多数のフレームを用いている。したがって、本方法の計算量は文献9)の方法より多い。例えば、入力画像の画素への参

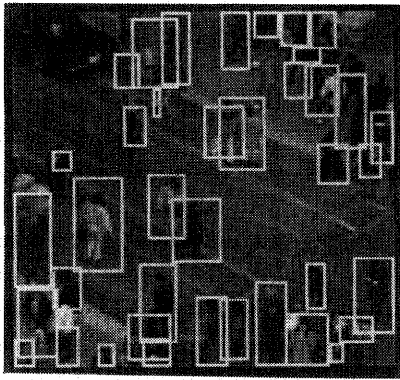


図 9 抽出された歩行者 (外接長方形で示す)
Fig. 9 Extracted pedestrians as circumscribed rectangles.

照回数で計算量を粗く評価すると、反復回数を 100 回とした文献 9) の方法に比べて、本方法 ($W=19$, ガウス関数の標準偏差 = 3 画素) の場合では参照回数はいはば 2 倍になる。

図 9 は、フレーム 100 において、本方法により歩行者を抽出した結果である。ただし、本方法のパラメータ $L_h=16$, $L_v=31$, 閾値 $\varepsilon_5=60^\circ$, $\varepsilon_6=1.0$, $\varepsilon_7=12$, $\varepsilon_8=14$ とした。外接長方形が抽出された歩行者の位置を表している。フレーム 101 から 120 においても同様の抽出処理を行い、合計 21 フレームにおいて抽出処理を行った結果、抽出率 86% (のべ歩行者数 819 人に対して正しく抽出されたのべ歩行者数 703 人) という結果を得た。主な誤りは、隣接する二人の歩行者が同じ動きをする場合、一人の歩行者として抽出されてしまうというものであった。

5. む す び

時空間輪郭積分による動柔物体の検出方法を提案した。本方法を、複雑な背景上に多数の歩行者が不均一な動きをしている横断歩道の動画像に適用して、カメラを 1.23 秒静止して約 86% の歩行者を抽出できた。

今後の課題として、動物体追跡への拡張¹⁴⁾、動物体の全輪郭の抽出、今回は実験的に決めた閾値の最適化がある。

謝辞 本研究の機会を与えていただいたマルチメディア処理研究部の遠藤隆也部長ならびに音声情報処理研究部の小林幸雄部長に感謝する。また、実験データを提供していただいた第 6 プロジェクトの塩 昭夫主幹研究員に感謝する。

参 考 文 献

- 1) Takagi, M. and Sakaue, K.: The Analysis of Moving Granules in a Pancreatic Cell by Digital Moving Image Processing, *Proc. 4th IJ CPR 1978*, pp. 735-739 (1978).
- 2) Takatoo, M. et al.: Traffic Flow Measuring System using Image Processing, *Proc. SPIE Vol. 1197 Automated Inspection and High-Speed Vision Architectures III 1989*, pp. 172-180 (1989).
- 3) Shio, A. and Sklansky, J.: Segmentation of People in Motion, *Proc. IEEE Workshop on Visual Motion 1991*, pp. 325-332 (1991).
- 4) 前田, 塩, 石井: 正規化主成分特徴量を利用した物体抽出法とその定量的評価, 電子情報通信学会論文誌, Vol. J75-D-II, No. 10, pp. 1660-1672 (1992).
- 5) Nagel, H.-H.: Formation of an Object Concept by Analysis of Systematic Time Variations in the Optically Perceptible Environment, *Comput. Gr. Image Process.*, Vol. 7, pp. 149-194 (1978).
- 6) Jain, R., Martin, W. N. and Aggarwal, J. K.: Segmentation through the Detection of Changes due to Motion, *Comput. Gr. Image Process.*, Vol. 11, pp. 13-34 (1979).
- 7) Suzuki, S. and Sklansky, J.: Extracting Non-rigid Moving Objects by Temporal Edges, *Proc. 11th IAPR International Conference on Pattern Recognition 1992, Vol. A*, pp. 69-73 (1992).
- 8) Duncan, J. H. and Chou, T.: On the Detection of Motion and the Computation of Optical Flow, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 14, No. 3, pp. 346-352 (1992).
- 9) Horn, B. K. P. and Schunck, B. G.: Determining Optical Flow, *Artificial Intelligence*, Vol. 17, pp. 185-203 (1981).
- 10) Burt, P. J. et al.: Object Tracking with a Moving Camera—An Application of Dynamic Motion Analysis, *Proc. IEEE Workshop on Visual Motion*, pp. 2-12 (1989).
- 11) Heeger, D. J.: Model for the Extraction of Image Flow, *J. Opt. Soc. Am. A*, Vol. 4, No. 8, pp. 1455-1471 (1987).
- 12) Fleet, D. J. and Jepson, A. D.: Computation of Component Image Velocity from Local Phase Information, *Intern. J. Computer Vision*, Vol. 5, No. 1, pp. 77-104 (1990).
- 13) 山本正信: 画像化された運動軌跡による動画像処理, 情報処理学会論文誌, Vol. 22, No. 5, pp. 442-449 (1981).

- 14) 鈴木, スクランスキー: Extracting and Tracking Nonrigid Moving Objects by Temporal Edges, 画像の認識・理解シンポジウム MIRU '92, pp. I-121-I-128 (1992).

(平成4年12月15日受付)

(平成5年6月17日採録)



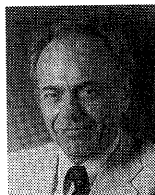
鈴木 智 (正会員)

昭和31年生。昭和54年静岡大学工学部情報工学科卒業。昭和56年同大学院修士課程修了。昭和59年同大学院博士課程修了。工学博士。同年、日本電信電話公社に入社。以来、線図形認識、地図データベース構築システムの研究開発に従事。平成3年から平成4年カリフォルニア大学アーバイン校客員研究員。現在、NTT ヒューマンインタフェース研究所研究企画部担当課長。コンピュータビジョン、画像認識・理解、画像処理アルゴリズム、並列処理に興味をもつ。電子情報通信学会、IEEE 各会員。



中根 一成

昭和22年生。昭和45年慶應義塾大学工学部電気工学科卒業。昭和47年同大学院修士課程修了。同年、日本電信電話公用(現、NTT)に入社。以来、図形情報システム、ビデオテックス用情報入力システム、ファクシミリ通信システム、会議通信システムなど主に画像通信システムの研究実用化に従事。現在、NTT ヒューマンインタフェース研究所視覚情報処理研究グループ・グループリーダー。電子情報通信学会、IEEE 各会員。



Jack Sklansky received the B. E. E. degree from the City College of New York in 1950, the M. S. E. E. degree from Purdue University in 1952, and the Doctor of Engineering Science degree from Columbia University in 1955. He is a Professor of Electrical and Computer Engineering, University of California, Irvine. His research has been concerned with digital pattern recognizers and trainable machines, with applications to digital video analysis and medical image understanding. He is an Editor-in-Chief of Machine Vision and Applications, on the advisory board of IEEE Trans. on Pattern Analysis and Machine Intelligence, and the editorial board of Pattern Recognition. He has published over 150 papers in his fields of interest.