

# アクティブチューブによる非剛体物体の抽出と追跡

## Non-Rigid Motion Tracking and Contour Extraction using Active Tubes

高畑 実                      今井 正和                      辻 三郎  
Minoru TAKAHATA    Masakazu IMAI    Saburo TSUJI

大阪大学 基礎工学部  
Faculty of Engineering Science, Osaka University

### Abstract

Analysis of non-rigid objects in image sequences is an important and difficult problem in computer vision. An input image sequence should be handled as a single image volume. We propose an energy minimizing deformable surface model, named Active Tube, for motion tracking and contour extraction in the spatio-temporal volume. We also try to extract initial positions for Active Tube automatically. The experimental results for synthetic images are presented.

### 1. はじめに

本報告では、複雑な背景下において運動する非剛体物体の輪郭を抽出する問題を扱う。

これは、環境の変化に対してロバスト、かつ、認識対象に依存しない汎用的な認識処理を目指すコンピュータビジョンの研究にとって基本的で重要な課題である。

非剛体の物体は一般的に特徴を捉えにくく、動きの情報が大変な手掛かりとなる。例えば、人間は運動情報のみから物体の輪郭を正確に知覚できることが知られており、心理学的知見に基づく運動物体の輪郭知覚モデル<sup>1)</sup>が提案されている。

動きによって対象を抽出する基礎的な研究には連続な差分画像を用いるものがある<sup>2), 3)</sup>。また、物体の速度が滑らかに変化するという制約を設けて時空間微分により画像のオプティカルフローを求める手法<sup>4)</sup>が提案されている。しかし、前者は物体の形状と運動方向が一定である必要があり、また、後者は物体表面の明度変化が滑らかである必要があつて、双方ともノイズに弱い。性能向上のために時空間微分と物体の特徴点の照合の両方を用いる手法<sup>5)</sup>も提案されている。

しかし、数多い手法が開発されているものの決定的なものはない。単純な背景をもつ画像から滑らかに運動する剛体性の高い物体の輪郭抽出には成功しているものの、複雑な背景をもつ画像からの非剛体物体の輪郭抽出はまだ十分な精度とは言い難い。

そこで、われわれは時空間画像を用いる。運動軌跡

を画像化する手法<sup>6), 7)</sup>が提案されているが、ここでは断面画像ではなく、入力される連続な画像列を重ねて作成される立体的な画像をそのまま用いる。このような時間的かつ空間的に連続な時空間画像を用いて対象の抽出を行えば、ノイズに強く、背景の影響等を受けにくい処理が期待できる。時間軸方向にも密な画像は対応点の決定問題を回避し、軌跡の交差を調べることでオクルージョンの問題を解決できる利点がある。

また、近年、最適化原理<sup>8)</sup>に基づく動的なモデルを用いる手法が数多く発表されている。

Snakes<sup>9)</sup>にみられるような輪郭抽出や領域抽出<sup>10)</sup>等の初期視覚問題に適用した例に始まり、物理法則に基づく形状モデルの構築法<sup>11), 12)</sup>が提案され、また、運動追跡についてはカルマンフィルタを用いた研究<sup>13), 14)</sup>、他、弾性を考慮したモデル<sup>15)</sup>などが報告され、その有効性が確認されている。

本報告では、エネルギー最小化の枠組みにより、静止した撮像系より入力された時空間画像中から、運動する非剛体物体の輪郭抽出と運動追跡を行なう動的なチューブのモデル**Active Tube**を提案する。このモデルで表現される物体の輪郭は時空間中でエネルギー最小となる立体表面として定義される。また、動的なモデルでは初期値が重要である。この部分は従来、人とのインタラクションなどが前提となっていた。初期値をエッジベースではなく、領域ベースで自動的に求める。これらの手法および基礎的な実験結果を示す。

## 2. 初期位置推定

エネルギー最小化の枠組により時空間画像における動領域表面の抽出を行うためにはなるべく初期位置を解の近傍に定めることが望ましい。そのような初期位置を選択することで **Active Tube** を望ましいエネルギー最小の位置へ収束させることができ、また、処理の速いアルゴリズムの選択を可能にする。カメラ静止の条件を仮定するので、具体的な処理手法は下図のようになる。

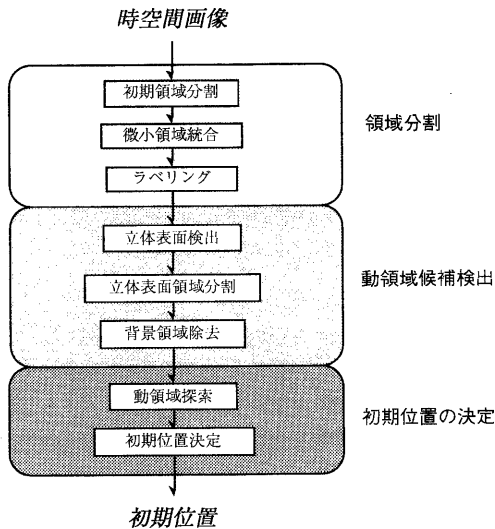


図1 初期位置推定アルゴリズム

まず、初めに時空間画像を領域分割する。そして得られる各々の立体形状を解析し、背景に対応する領域を取り除く。その後、動領域中で物体に速度制約を課すフィルタをかけ、初期位置を決める。

### 2.1 領域分割

初期領域分割ではヒストグラム分割と領域成長法により時空間画像をセグメンテーションする。

ヒストグラム分割は、空間座標一定で時間軸方向の明度ヒストグラムをとり、図2のように前処理としてヒストグラム空間上でガウシアンフィルタを施した後、谷-谷で閾値処理を行い、各クラスの代表値（モード）をそれぞれのクラスに埋めることで分割する。

領域成長法はある画素を考えたとき、その近傍の画素があらかじめ決めておいた閾値以下ならばその画素を融合し、これを繰り返す手法である。

具体的には、時間領域においてヒストグラムを用いた領域分割を行った後、空間領域で領域成長法による領域分割を行うという2段階の構成によりノイズの影響を抑制する。

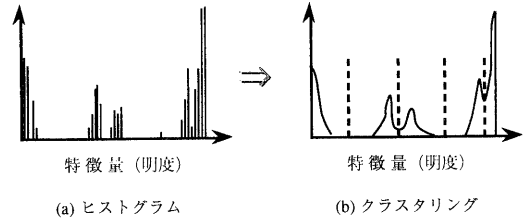


図2 ヒストグラム領域分割

### 2.2 動領域候補検出

立体表面領域分割では背景に対応する領域を探索し除去する。時間軸に平行な表面は背景領域間のエッジであるので、まず、't-constant'のcriteriaによって時間軸に平行な表面とそうでない表面に各々の立体表面をセグメンテーションする。

そして、そのようなエッジで囲まれている領域は背景であるとして除去する。

#### \* 't-constant' criteria

$$\exists P = (x_1, y_1, t_1) \subseteq S \text{ (立体表面) の近傍で}$$

$$\sum_{t=t_1 - \text{SENSOR}}^{t_1} P(x_1, y_1, t) / \text{SENSOR} \geq \text{Threshold}$$

または、

$$\sum_{t=t_1}^{t_1 + \text{SENSOR}} P(x_1, y_1, t) / \text{SENSOR} \geq \text{Threshold}$$

ならば 静止 voxel

### 2.3 初期位置の決定

初期位置の決定では動領域の候補を更に探索する。動領域候補検出の段階で残った領域で、静止していると見なせる部分を除去する（図3斜線部）。

また、領域を切断することにより生じる速度無限大の不連続面（図4斜線部）を速度制約を課すフィルタによって削除する。ある時間での切断面は、その前後の切断面と比べて大幅な変化がない性質を利用し、領域の膨張と比較操作により実現している。

最終的に図5のような初期位置を得る。

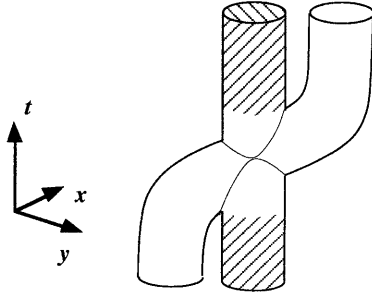


図3 動領域の候補の例 (斜線部は静止領域)

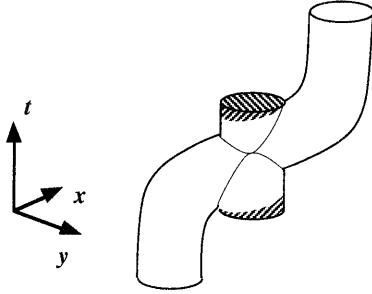


図4 速度制約フィルタ (斜線部は速度の大きな領域)

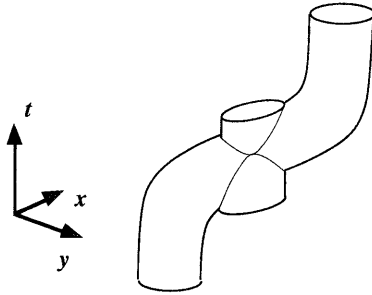


図5 初期位置の例

### 3. Active Tube

動的なチューブのモデルは、動物体の輪郭から生成される時空間画像中の立体表面へのパラメータ  $s, t$  によるベクトル  $v(s, t)$  で表現される。そして、以下の式 (1) のエネルギー汎関数を最小にする変関数  $v(s, t)$  が求める領域の表面である。パラメータ  $s$  は空間領域に、パラメータ  $t$  は時間領域における成分であり、それぞれ 2 次元と 1 次元の空間を表現できる。すなわち、ここで提案する **Active Tube** は 2 次元の空間に、性質の異なる時間の

次元をひとつ加えた 3 次元上での 2 次元的な立体表面のモデルである。

また、**Active Tube** はエネルギー最小の意味で最適な、時空間中の立体表面抽出のための動的なモデルである。

$$E_{\text{Tube}}^* = \int_T \int_0^1 E_{\text{int}}(v(s, t)) + E_{\text{image}}(v(s, t)) ds dt \quad \dots (1)$$

式 (1) のエネルギー汎関数は、立体表面の曲率と連続性に依存する内部エネルギー  $E_{\text{int}}$  と、時空間画像のグラディエントから生じるポテンシャルに依存するエネルギー  $E_{\text{image}}$  からなる。

$$E_{\text{int}} = \alpha(s, t) |v_s(s, t)|^2 + \beta(s, t) |v_{ss}(s, t)|^2 + \gamma(s, t) |v_{tt}(s, t)|^2 \quad \dots (2)$$

内部エネルギーは式 (2) のように表現される関数である。空間領域での一次微分項  $\alpha(s, t) |v_s(s, t)|^2$  は空間方向の連続性に、空間領域での二次微分項  $\beta(s, t) |v_{ss}(s, t)|^2$  は空間方向の滑らかさに、また、時間領域での二次微分項  $\gamma(s, t) |v_{tt}(s, t)|^2$  は時間方向の滑らかさにそれぞれ依存する。 $\alpha, \beta, \gamma$  はパラメータであり、 $E_{\text{int}}$  は  $\alpha, \beta, \gamma$  が各々正の値を持つ場合、立体表面を表現するベクトルがなるべく均一で滑らかになるようとする力を持つ。

なお、ここで提案するモデルは時間軸方向に密であるため、時間軸方向の一次微分項は持たない。

$$E_{\text{image}} = w_{\text{edge}} E_{\text{edge}} + w_{\text{t-smooth}} E_{\text{t-smooth}} + w_{\text{feature}} E_{\text{feature}} \quad \dots (3)$$

画像からのエネルギーは式 (3) のように表現される関数であり、画像の特徴に引きつけたり、遠ざけたりする作用を持つ。

ここで、

$$E_{\text{edge}} = - \left| \frac{\partial}{\partial x} I(x, y, t) + \frac{\partial}{\partial y} I(x, y, t) \right|^2 \quad \dots (4)$$

$$E_{\text{t-smooth}} = \left| \frac{\partial}{\partial t} I(x, y, t) \right| \quad \dots (5)$$

と定義すれば、 $E_{\text{edge}}$  は空間方向のエッジに、 $E_{\text{t-smooth}}$  は時間方向のエッジにそれぞれ依存する。各々のパラメータ  $w$  が正の値を持つ場合、 $E_{\text{image}}$  は空間方向のエッジがある方向への力を持ち、しかも時間方向のエッジ、つまり静止しているエッジを避けようとする方向の力も持つ。また、もし  $E_{\text{feature}}$  として空間領域での 2 次微分項を考えれば、高い曲率値をもつ部分に引きつけることになり、特徴の追跡に役立つ。

### 3.1 アルゴリズム

**Active Tube** のエネルギーは汎関数であるので、最小化の必要条件はオイラー・ラグランジュの方程式となり、汎関数 (1) を極小にする変関数  $v(s, t)$  を求めるには、これを反復法で解く。すなわち、変分法により解くことが考えられるが、安定性や収束性等の問題があってパラメータの設定を困難にしている。

この問題を解決するために動的計画法 (DP) による解法<sup>16)</sup>があるが、速度の点で問題となる。離散化表現された立体表面の点の総数を  $n$  点、各点の近傍を  $m$  点と考えると計算の複雑さは  $O(nm^3)$  になる。また、閉曲線に対し厳密に定式化すれば計算の複雑さは  $O(nm^5)$  にも達することが報告されている<sup>17)</sup>。しかも、提案する時空間上のモデルでは点の総数  $n$  が必然的に大きくなり、より緻密な絵では画像サイズの増大と共に現実味が薄れる。

そこで、Greedy アルゴリズム<sup>18)</sup>を時空間に拡張して用いた。Greedy アルゴリズムは輪郭線を表す点をその近傍でエネルギー最小となるところへ順次に動かし、局的最小となる所へ輪郭線を収束させる解法である。この方法によれば計算の複雑さは  $O(nm)$  になる。ただ、この方法では初期位置から大きく変化して動くような場合に必ずしも良好な解を得られるとは限らない。しかし、この問題は先に述べた初期位置推定法によって収束目標になるべく近い初期位置を求めることで解決できる。

今、 $v(s, t) = v_{i,t}$  ( $i = 1, \dots, n, t = 1, \dots, m$ ) とするとき、**Active Tube** のエネルギーは、

$$E_{\text{Tube}} = \sum_{t=1}^m \sum_{i=1}^n E_{\text{int}}(v_{i,t}) + E_{\text{image}}(v_{i,t}) \quad \dots (6)$$

右辺第一項目の内部エネルギーは、

$$E_{\text{int}}(v_{i,t}) = \alpha_{i,t} \left| \frac{\partial v_{i,t}}{\partial s} \right|^2 + \beta_{i,t} \left| \frac{\partial^2 v_{i,t}}{\partial s^2} \right|^2 + \gamma_{i,t} \left| \frac{\partial^2 v_{i,t}}{\partial t^2} \right|^2$$

空間領域では、なるべく等間隔に点が並ぼうとするので、

$$\left| \frac{\partial v_{i,t}}{\partial s} \right|^2 \approx |v_{i,t} - v_{i-1,t}|^2$$

$$\left| \frac{\partial^2 v_{i,t}}{\partial s^2} \right|^2 \approx |v_{i-1,t} - 2v_{i,t} + v_{i+1,t}|^2$$

と近似できる。時間領域においては各点は時間軸方向に動かず、等間隔にならない。そこで図6のように、

$$V_1 = v_{p,t-1} - v_{i,t}, \quad V_2 = v_{q,t+1} - v_{i,t}$$

$v_{p,t-1}$  は立体表面の点で時刻  $t-1$  において  $v_{i,t}$  に最も近い点

$v_{q,t+1}$  は立体表面の点で時刻  $t+1$  において  $v_{i,t}$  に最も近い点

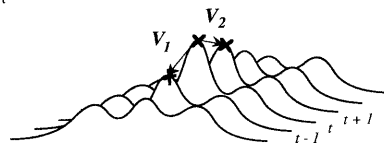


図6 時空間立体表面

とするとき、

$$\left| \frac{\partial^2 v_{i,t}}{\partial t^2} \right|^2 = \cos^{-1} \left( \frac{V_1 \cdot V_2}{|V_1| |V_2|} \right)$$

となる。

式 (6) の右辺第二項目の画像によるエネルギーは2章の式 (3), (4), (5) で示した。

また、パラメータの設定を簡素化するため、各々のエネルギーは各点及びその近傍で計算した後、正規化され、0~1の値を持つ。

### 4. 実験結果

まず始めに合成された画像列を用いて実験を行った。入力画像列は複雑な背景の中を、変形と同時に上下に振動しながら横方向へ運動する楕円を捉えたものである。

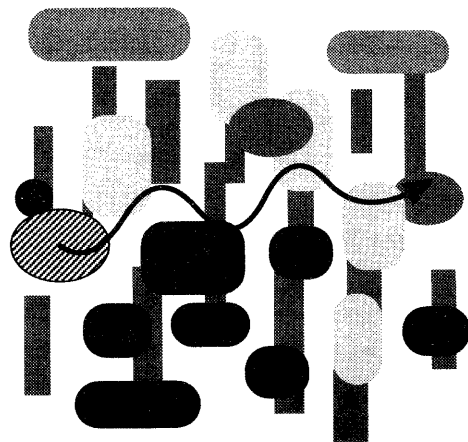


図7 複雑背景下を移動する楕円

背景は32個の幾何学的な図形によって構成されており、移動する楕円は背景中の15個の長方形と明度、彩度及び色相が同一である。また、画像列全体に標準偏差4の白色ガウシアンノイズを加えた。なお、画像のサイズは64×64である。このような画像列を64枚重ねて、一つの時空間画像を作成した。一部をカットした様子を下図に示す。

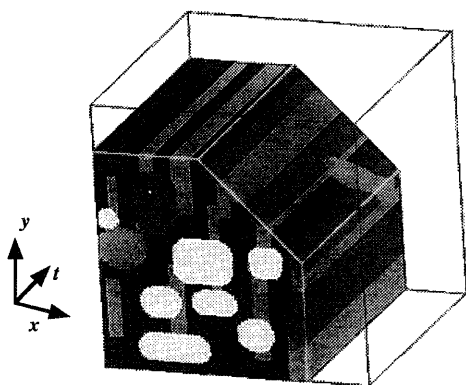


図8 時空間画像

提案するアルゴリズムはC言語によりSun 4/75 Workstation上にインプリメントし、実験を行った。

まず、初期位置推定ではヒストグラムによる分割のためのガウシアンフィルタの $\sigma$ を1.0、領域成長法の閾値を24、時間軸に平行なエッジを見つけるためのセンサー長を4[frame]、また、速度制約フィルタの制限値を3[pixel/frame]とし、実験を行った。得られた初期位置の時空間立体表面にレンダリングを施し、その結果を図9に示す。このとき、初期位置推定の全行程で約10分を要した。計算時間は領域の数に比例する。

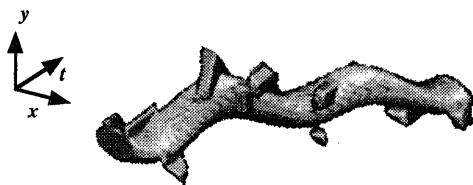


図9 合成画像より推定された初期位置

次に得られた初期位置のうち、2,165点を使用しActive Tubeの実験を行った。ここでは、パラメータは定数であって、それぞれ $\alpha = 1.0$ ,  $\beta = 1.0$ ,  $\gamma = 2.8$ ,  $W_{\text{edge}} = 1.4$ ,  $W_{t\text{-smooth}} = 1.0$ を用いた。

現段階ではEfeatureはインプリメントしていない。

実験では繰り返し30回目にほぼ収束した。そのときの収束途中の様子を図12、13に、また、収束した結果を図14に示す。図7斜線部の楕円の輪郭抽出及び運動追跡を行えたことが確認された。このとき、繰り返し一回につき1秒強を要している。

なお、図8、図9、図12～14(c)は、おおよそ同じような視点から眺めた図となっている。

次に、実画像による実験を行った。入力画像列は静止した撮像系により、ジョギングマシン(treadmill)の上で走る人を捉えたものである。下図は入力画像列の中の1フレームである。

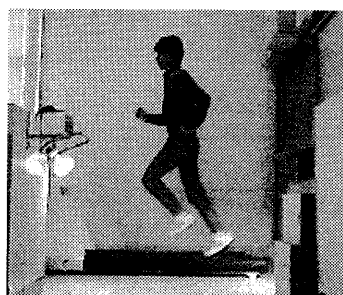


図10 走行する人

領域成長法の閾値を16にした以外は、先の実験と条件を揃えて実験を行った。同様に、得られた初期位置の時空間立体表面にレンダリングを施し、その結果を下図に示す。計算時間は約15分を要した。

得られた初期位置は、右下(足元)の部分で陰があり、同一時間に領域が複数存在する。現在、そのような場合にActive Tubeは対応できないので、今後の課題である。

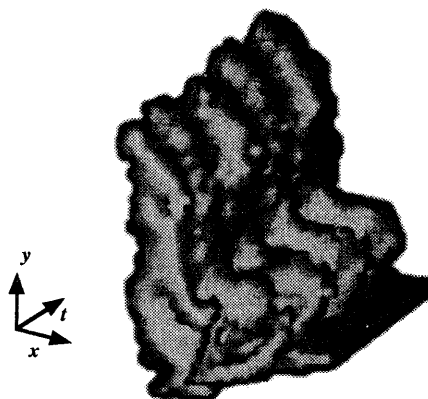


図11 実画像より推定された初期位置

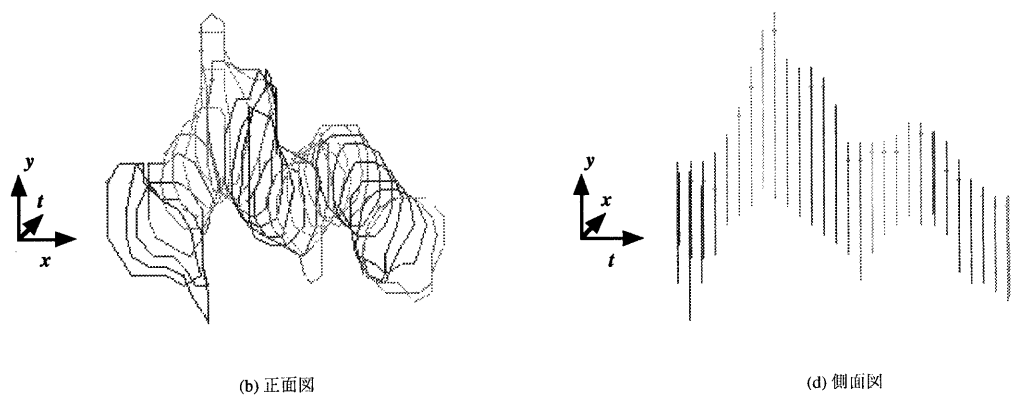
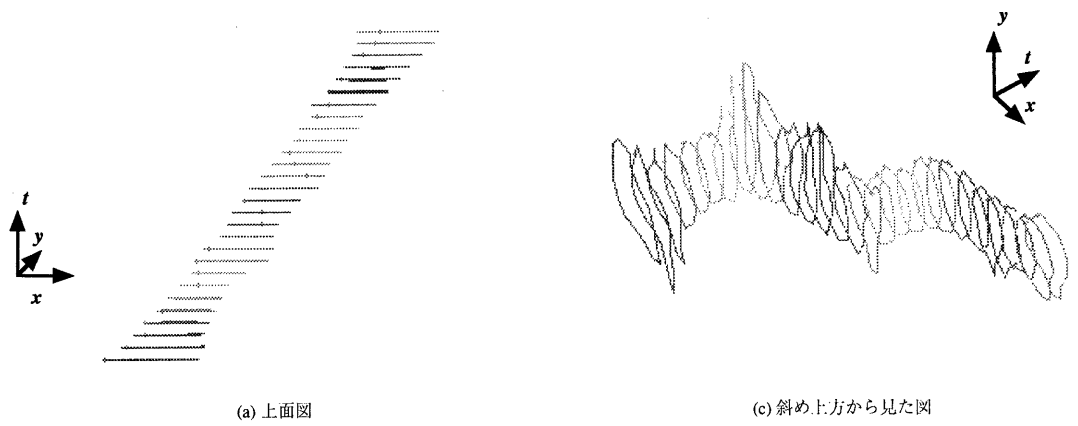


図 12 途中結果（繰り返し1回目）

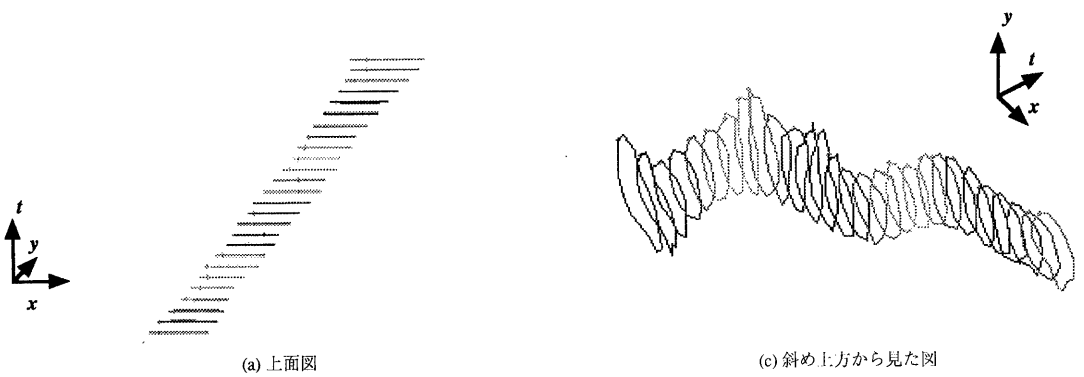
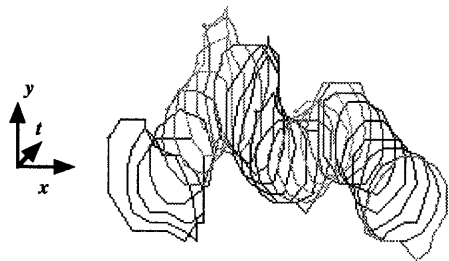
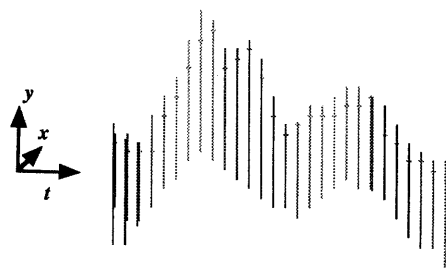


図 13 途中結果（繰り返し8回目）  
（次のページへ続く）

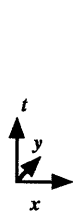


(b) 正面図

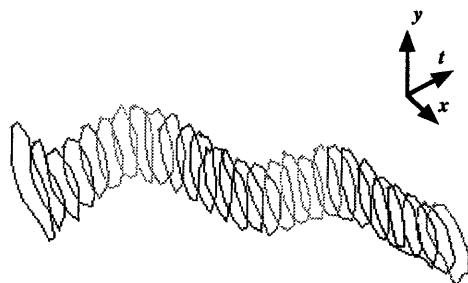


(d) 側面図

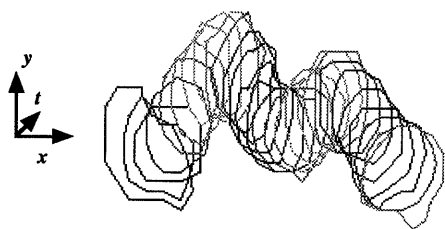
図 1 3 (前のページからの続き) 途中結果 (繰り返し 8 回目)



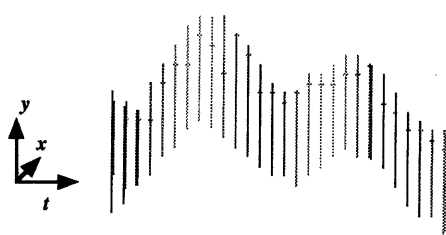
(a) 上面図



(c) 斜め上方から見た図



(b) 正面図



(d) 側面図

図 1 4 収束した結果 (繰り返し 3 0 回目)

## 5. おわりに

時空間画像から動物体の輪郭の抽出と運動の追跡を行う動的なモデル**Active Tube**とその初期位置推定法について述べた。まず、初期位置推定の段階でおおよそのセグメンテーションが行われ、続いて**Active Tube**がその結果をリファインする形になっている。**Active Tube**の収束とは、輪郭抽出と運動追跡の完了と同義である。

また、**Active Tube**は対象物体のモデルを持っていないため、柔物体など広範囲に適用可能である。

さらに、提案する手法はある程度長い時間の中で考えるため、つまり強力な時間制約を用いるために空間方向のノイズに対してロバストであり、仮に動物体が背景の一部と全くおなじであっても有効である。しかし、初期位置推定の領域分割の段階において、時間方向のノイズなどの原因であまりにも多数の領域が生じる場合には弱い。

今後の課題には、2つ以上の物体の存在や陰などの影響によって、複数の動領域が重なる場合の対処法がある。ほかに、初期位置推定の段階で、どこが背景として削除されたかをパラメータ設定に利用することも考えられる。

**Active Tube**のようなエネルギー最小化原理に基づく考え方は、より高いレベルからのフィードバックができることに大きな意義があって、従来の画像処理のように各段階で決定的な方法とは一線を画す手法であると考えており、今後さらに発展させる計画である。

## 【参考文献】

- 1) 緒方, 佐藤: "2段階処理による運動物体の形状知覚モデル", 信学技報, PRU 89 - 44 (1989)
- 2) M.Yachida, M.Asada and S.Tsuji: "Automatic Analysis of Moving Images", IEEE Trans. on PAMI, Vol.3, No.1, pp.12 - 20 (1981)
- 3) R.Jain and H.H.Nagel: "On the Analysis of Accumulative Difference Pictures from Image Sequence of Real World Scenes", IEEE Trans. on PAMI, Vol.1, pp.206 - 214 (1979)
- 4) B.K.P.Horn and B.G.Schunck: "Determining Optical Flow", Artificial Intelligence, Vol.17, pp.185 - 203 (1981)
- 5) M.Yachida: "Determining Velocity Maps by Spatio-Temporal Neighborhoods from Image Sequence", CGIP, Vol.21, pp.262 - 279 (1983)
- 6) 山本: "画像化された運動軌跡による動画像処理", 情処学論, Vol.22, No.5, pp.442 - 449 (1981)
- 7) R.C.Bolles, H.H.Baker and D.H.Marimont: "Epipolar-Plane Image Analysis: An Approach to Determining Structure from Motion", International Journal of Computer Vision, 1, pp.7 - 55 (1987)
- 8) 坂上: "画像処理の研究動向", テレビジョン学会誌, Vol.45, No.10, pp.1147 - 1154 (1991)
- 9) M.Kass, A.Witkin and D.Terzopoulos: "Snakes: Active Contour Models", International Journal of Computer Vision, Vol.1, No.4, pp.321 - 331 (1988)
- 10) 積山, 坂上, 山本: "Active Net: 動的な網のモデル", 信学技報, CV 63 - 2 (1989)
- 11) D.Terzopoulos and D.Metaxas: "Dynamic 3D Models with Local and Global Deformations: Deformable Superquadrics", IEEE Trans. on PAMI, Vol.13, No.7, pp.703 - 714 (1991)
- 12) A.Pentland and S.Sclaroff: "Closed-Form Solution for Physically Based Shape Modeling and Recognition", IEEE Trans. on PAMI, Vol.13, No.7, pp.715 - 729 (1991)
- 13) A.Pentland and B.Horowitz: "Recovery of Nonrigid Motion and Structure", IEEE Trans. on PAMI, Vol.13, No.7, pp.730 - 742 (1991)
- 14) D.Metaxas and D.Terzopoulos: "Recursive Estimation of Shape and Nonrigid Motion", Workshop on VISUAL MOTION, pp.306 - 311 (1991)
- 15) 上田, 間瀬, 末永: "エネルギー最小化に基づく輪郭追跡", 信学技報, CV 73 - 2 (1991)
- 16) A.A.Amini, T.E.Weymouth: "Using dynamic programming for solving variational problems in vision", IEEE Trans. on PAMI, Vol.12, No.9 (1990)
- 17) 上田, 間瀬: "動的計画法によるactive contour エネルギーの最小化", 1991年春季信学全大, D-554
- 18) D.J.Williams and M.Shah: "A Fast Algorithm for Active Contours", In Proc. of Third International Conference on Computer Vision, pp.592 - 595 (1990)