

## 映像シーン予測法

Video Scene Prediction Method

境野 英朋<sup>†</sup>Hidetomo Sakaino<sup>†</sup>

## 1. まえがき

コンピュータビジョンでは、時系列画像中の移動する物体(対象)の動きを検出し、予測する研究が盛んに行われている。その多くは、対象の特徴点や輪郭線の追従である。動きの予測は、ワイヤフレームモデル等の対象への当てはめ誤差を改善する手段としても利用されている。MPEGの動き補償では1フレーム先をblockマッチング法による予測が用いられている。ブロック歪に代表される画質の不自然さが指摘されているが、対象の輪郭線近傍を小さいブロックに分割することによる改善も試みられている。一般に、速度ベクトル推定改善等に用いられるセグメンテーションは対象の輪郭線が不明瞭な流体パターンの場合には適用が困難となる。また、テクスチャ変化と移動を同時に予測する研究もほとんどない。即ち、対象を含んだ画像フレーム(シーン)全体を領域分割することなく、ブロック単位から画素単位に予測できる手法が望まれる。そこで本稿では、ダイナミクス・テクスチャ法の適用を提案し、その有効性について論じる。

## 2. 画素単位の変形・移動予測方法

## 2.1 従来移動モデル

一般に、物体が移動することを予測する場合、どれだけの事前情報が得られるかで予測モデルが選択される。画像内の移動体のモデルについては、ここでは図1に示すような4つの場合について述べる。

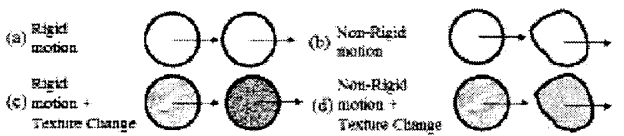


図1 対象の時系列変化モデル

まず、簡単な例として、画像中を水平方向に、球が飛んでいく場合があるとする。剛体(図1a)であれば、その重心を検出して速度や加速度を推定できれば、式(1)に示すようなオイラー形式で重心の予測を現在と過去情報からできる。ここで、 $\mathbf{x}(t)$ ,  $\mathbf{v}(t)$ は時刻  $t$  における対象の重心の2次元位置と速度である。加速度をゼロとした線形モデルである。

$$\mathbf{x}(t+1) = \mathbf{x}(t) + \mathbf{v}(t)t \quad (1)$$

次に対象が弾性体のように、変形を伴いながら移動していく場合を想定する(図1b)。この場合、コンピュータビジョン(CV)では動的輪郭線モデル[1]や動的メッシュ[2]をもつ超関数モデルをベースにした輪郭線の追従が適用されることが多い。本モデルでは弾性係数を事前に与える必要があるのと、対象の輪郭線が未来でどのようになっていくかの予測変形は、非周期性の場合、別のモデルが必要である。

対象の外形(輪郭線)の変形に加えて、実物はテクスチャ変化も伴いながら移動する場合も多い(図1c, d)。球が変形と移動を同時に伴った場合はどのように予測するかについてはほとんど議論されていない[3]。心臓や胃[1]などの診断画像の動的な輪郭追従や輪郭線の予測が見られるがテクスチャ変化モデルまでは示されていない[1][2][3]。テク

スチャをコンピュータグラフィックス(CG)で生成することが盛んに行われている。その際、変形は人為的に制御することでテクスチャの変形を実現している。CVでは、実写からモデルパラメータを推定する研究[4]が多いが動的なテクスチャについてはほとんどなされていない。対象の特徴点の追従[5]やカルマンフィルタによる予測等は多く見られるが、テクスチャ変形予測まではほとんどなされていない。上述したように、CVでは対象の属性情報についての事前知識を必要としており、モデルも対象依存である。また、対象のテクスチャ変形を伴った変形予測手法については十分に議論されていない。

## 2.2 非定常・非線形モデル

従来から画像パターンでは対象の特徴点や重心など、点ベースの予測方法と言える。図2aに示すように、過去3点を検出できれば未来の加速度変化を予測することができる。テクスチャ変化については別の処理が必要である。しかしながら、2つの時間情報しか得られない場合は事前知識がないと速度変化(加速)予測はできない。ここで、画像パターンが画素で構成されていることから、1画素ずつに速度をもたせて、画素単位の変化予測を考える(図2b)。各画素が対象の特徴を表す濃淡値(強度)をもち、テクスチャを形成する。各画素が速度の変化とともに、時刻  $t-1$  と  $t$  から未来のフレームを予測するとする。各画素がそれぞれの速度で移動し、同時に近傍の速度場が影響し合えばさまざまな速度場が生じる。結果として、非線形な速度場とテクスチャ変化を両立させられる。

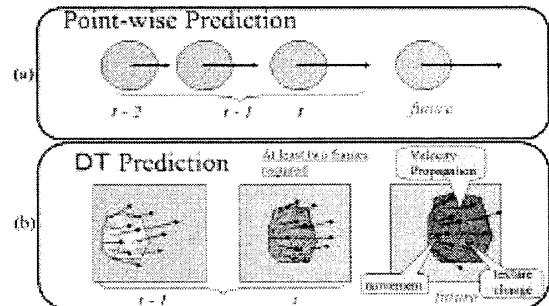


図2 点予測とDT法での面予測の相違

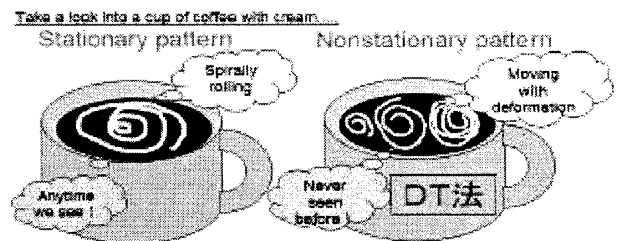


図3 クリームパターン変化と非定常性

予測ではパターンの非定常性も考慮する必要がある。図3に示すような通常観るコーヒーカップの中のクリームを例にとると、渦巻きパターンは定常パターンである。渦パターンが移動しながら変形するような2度と観ることができないようなパターンは非定常性と考えることができる。

<sup>†</sup> 日本電信電話(株), NTT コミュニケーション科学基礎研究所

複雑なパターンの濃淡値(テクスチャ変化)と移動を同時に行うためには、1)時系列フレーム間で点対応づけが必要、2)非線形で非定常予測であること、3)形状や属性に関する知識をほとんど使わずに済むこと、4)画素単位予測、5)速度場の伝播等が必要となる。そこで、これらの要件を満たすダイナミクス・テクスチャ(DT)法[6]に着目する。本手法は降水レーダパターン予測問題に既に適用されているが映像シーンへの適用については報告されていない。

DT 法の特徴は、1)移動速度(オプティカルフロー)をパターン情報から推定することで点対応づけを行っている、2)移流、拡散、発達、衰退、停滞等の基本変化を網羅した非線形・非定常予測を使用している、3)過去と現在の 2 フレームを使用する、4)移動平均フィルタで速度伝播、等である。一方、MPEG 等では、1)線形手法であること、2)速度伝播予測ができない、3)サブブロック同士の関係は独立、4)フレーム間での、発達、融合、分裂表現に乏しい、等であるので本予測問題には適用できない。

### 3. ダイナミクス・テクスチャ法による予測実験

本章では DT 法の概略を述べる。パターンの画像強度  $I$ 、位置  $\mathbf{x}$ 、時間  $t$  とする。空間  $(i, j)$ 、時間  $n$  の各微分値は計算格子点で差分近似した。式(2)の時間項を前進差分をとった式(3)を所定未来時間先まで反復的に時間幅  $\Delta t$  の間隔で繰り返す。STE $_i$  ( $i=1, 2, 3$ ) はそれぞれ移流項、異方拡散項、状態項である。非線形な移流項中、パターン速度  $\mathbf{u}$  は標準正則化法で速度に 1 次微分についての拘束条件を与えて求めたオプティカルフローにより与えた。また、速度にノイズ軽減のために特異値分解を適用した。速度場の予測は移動平均フィルタを反復的に適用した。状態項は発達・衰退・停滞に関して、時間差分値にしきい値により 3 つに分類した結果を与える。異方拡散項はエッジ構造を保ちつつ平滑化の効果をする。予測に際して、それぞれのレーダパターンの現在(初期値)と 1 つ過去の、連続する 2 フレームのみを使用した。low-pass filter 等の前処理は行っていない。 $h_x=h_y=1.0$ 。

$$\frac{\partial I(\mathbf{x}, t)}{\partial t} = \sum_{q=1}^3 STE_q, \quad (2)$$

$$I_{i,j}^{n+1} = I_{i,j}^n + \Delta t \cdot \sum_{q=1}^3 STE_q. \quad (3)$$

where

$$\begin{cases} STE_1 = Advect(I, \mathbf{u}, i, j, n), & \mathbf{x}_{i,j} = (x, y) = (h_x i, h_y j), \mathbf{u}_{i,j} = (u_x, v_x) \\ STE_2 = Diff(I, i, j, n), & \nabla \equiv \frac{\partial}{\partial x} + \frac{\partial}{\partial y} \\ STE_3 = State(I, i, j, n). \end{cases}$$

$$\begin{aligned} Advect(I, \mathbf{u}, i, j, n) &= \mathbf{u}_{i,j} \cdot \nabla I_{i,j} & Diff(I, i, j, n) &= I_{i,j}^{n+1} - I_{i,j}^n + \lambda \cdot \sum_{k=1}^4 [c_k^n \nabla_k I] \\ \nabla_1^n I &= (I_{i-1,j}^n - I_{i,j}^n) / \Delta x, \nabla_2^n I &= (I_{i+1,j}^n - I_{i,j}^n) / \Delta x, & g_k^{-1}(z) &= 1 + \frac{|z|^2}{k} \\ \nabla_3^n I &= (I_{i,j-1}^n - I_{i,j}^n) / \Delta y, \nabla_4^n I &= (I_{i,j+1}^n - I_{i,j}^n) / \Delta y, & c_k &= g \left( \frac{|\nabla_k^n I|}{k} \right) \end{aligned}$$

$$State(I, i, j, n) = I(i, j, n) / \Delta t - I(i, j, n-1) / \Delta t.$$

映像シーンとしては、輪郭線検出や領域分割が予備実験で困難であったものを選択し、数フレーム先を予測した。各図において、DT 法で予測したシーンには推定されたオプティカルフローベクトル(橙色)を重ねて描いてある。図4は超音波で撮った心臓の収縮であり、紡錘状の心臓のまわりには内膜があり、心臓本来の境界線は非常に曖昧である。DT 法では濃淡値全体と形状が 6 フレーム先でも良好に予測生成できた。図5は会話をしているときに、ある実験で器

具を付けた表情変化であり、唇と皮膚の境界が曖昧であるが、DT 法では 9 フレーム先まで剛体や弾性体が混合しているにも関わらず的確にシーン全体の予測ができている。図6は北極オーロラのカーテン状の変化である。シーン下に森林の影がある。オーロラは濃淡値変化や発達・衰退を伴いながら移動したが、DT 法では波状に広がった。いずれのシーンでも従来ブロック予測では濃淡値変化がなく、分裂状にもなった。ここには示していないが、他の予測法に依存する場合も、シーン毎に特定の形状モデルや輪郭線検出、領域分割等が必要である。

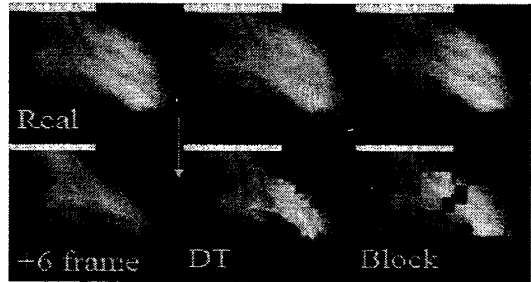


図4 心臓収縮過程での映像予測の比較

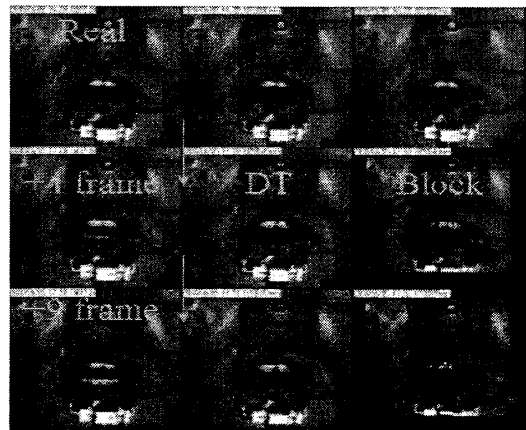


図5 器具をつけた状態での表情(会話)での映像予測

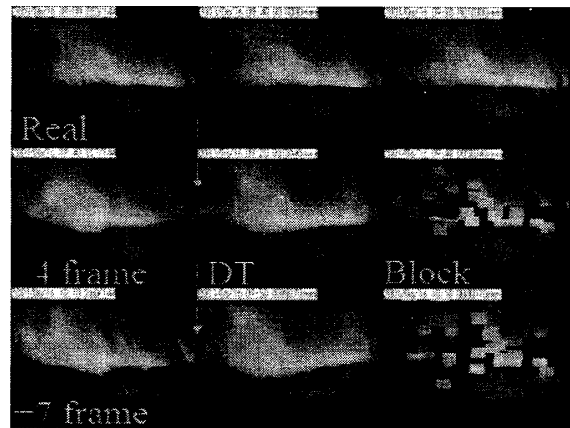


図6 北極オーロラのカーテン状のパターン変化予測

### 4.まとめ

人の視空間でのシーン予測のメカニズムを探索していく。  
参考文献 [1]Y.Kita, "Elastic...", PAMI, 96. [2]D.DeCarlo et al., "Deformable models", PAMI, 96. [3]Y.Yacoub et al., "Tracking rigid...", ICCV, 99. [4]E.J.Pauwels et al., "Cluster...", ICCV 99. [5]N.Paragios et al., "Geodesic...", ICCV, 99. [6]境野, 信学会 IE2002-5, pp43-48, 02.