# 動画像における時空間ダイナミクスのモデル化

## 川嶋 宏彰<sup>†</sup> 三井 $dd^{\dagger}$ 松山 隆司<sup>†</sup>

#### †京都大学大学院情報学研究科 〒606-8501 京都市左京区吉田本町

E-mail: *†*{kawashima,tm}@i.kyoto-u.ac.jp, *†*†mitsui@vision.kuee.kyoto-u.ac.jp

あらまし 動画像におけるダイナミクスとして,従来,2次元信号の時間変化(テクスチャダイナミクス),および 注目対象領域の変形や運動(領域ダイナミクス)の2つが扱われてきた.前者は,波立つ水面領域のように対象領域 が固定された場合に,後者は,歩行者追跡などで対象の輪郭変化を表現したい場合に,それぞれ有効なモデルとなる. しかし,風になびく大木の葉の映像のように,動的に変化する領域とその領域内のテクスチャ変化とを同時にモデル 化しようとすると,いずれか一方のモデル化だけでは限界がある.そこで本研究では,これら2つのダイナミクスを 結びつけるひとつの枠組みとして,領域ダイナミクスに基づいて局所的なテクスチャダイナミクスの切り替わりをモ デル化する新たな手法を提案する.

キーワード ダイナミックテクスチャ、領域ダイナミクス、レベルセット関数,動的システム、時空間領域分割

# Modeling Spatio-Temporal Dynamics in Video Images

## Hiroaki KAWASHIMA<sup>†</sup>, Ken MITSUI<sup>†</sup>, and Takashi MATSUYAMA<sup>†</sup>

† Graduate School of Informatics, Kyoto University Yoshida-Honmachi, Sakyo, Kyoto, 606-8501 Japan E-mail: †{kawashima,tm}@i.kyoto-u.ac.jp, ††mitsui@vision.kuee.kyoto-u.ac.jp

**Abstract** Modeling dynamics in video images is crucial for many applications including generation of natural scenic images and construction of dynamic background models for object detection in non-stationary scenes. Dynamics in video images is categorized into two types: texture dynamics (i.e., the time variation of two-dimensional signals in a certain area) and region dynamics (i.e., the deformation or motion of the object that has its shape). Prior studies proposed various methods for modeling dynamic scene in video images by using either one of two dynamics. However, those methods have limitations in order to model the time variation of texture in a deforming region. In this paper, we propose a method to simultaneously model texture dynamics and region dynamics. **Key words** Dynamic texture, shape dynamics, level-set function, dynamical system, spatio-temporal segmentation

### 1. はじめに

風景は,光と影により浮かび上がるパターン(景)だけでな く,その時間変化(風,いわばダイナミクス)とが織り成され た持続的な様子である.したがって,自然な風景映像の生成や, 移動対象検出のための背景モデル構築を行うには,風景を撮影 して得られた動画像のダイナミクス,すなわち木の葉や水面, 空などの様々な時間変化を表現することが重要となる.

この動画像におけるダイナミクスには,大きく分けて以下の 2つのものが考えられ,従来,それぞれをモデル化するための 手法が提案されてきた.

テクスチャダイナミクス 画像全体もしくは画像中の部分領域 におけるテクスチャパターンの時間変化.これは2次元信号と してのダイナミクスである. 対象領域のダイナミクス 画像中に存在する対象物体の動き. これは移動や回転,スケール変化といった幾何学的変換と,形 状そのものの幾何学的変形があり,しばしば分けて扱われる.

テクスチャダイナミクスのモデル化手法としては,波立つ水 面領域のように,ダイナミクスを記述する対象領域が固定され た場合に,その領域内におけるテクスチャ変化を,動的モデル として表現するものがある.例えば,Dynamic Texture [1] は, 画像全体における画素の輝度値の時間変化を線形システムで記 述する手法であり,そのシンプルさから,動的背景や映像生成 への応用が近年盛んに行われている [2],[3].これらの手法では, モデル化単位である画像全体において,テクスチャダイナミク スが時間的に一定であることを仮定しているが,画像中に水面 や空などが含まれる場合,それぞれのテクスチャが異なる速度 やパターンで変化する.このような動画像では,画像全体では



図 2 テクスチャダイナミクスと領域ダイナミクスの統合

**Object-region Dynamics** 

なく,画像中の部分領域をモデル化単位とし,その部分領域に おいてテクスチャダイナミクスが時間的に一定であると仮定し て,そのモデル化を行うという手法が用いられる[4].

しかし,部分領域によってはテクスチャダイナミクスが時間 軸上で一定とはいえない場合がある.例えば,図1に示す矩形 小領域では,時間によって木の葉と壁面という異なる対象が交 互に現れており,異なるテクスチャパターンおよびその時間変 化が観察される.したがって,このような小領域におけるテク スチャダイナミクスを,単一のモデルで精度良く表現すること は困難であり,対象領域を考慮したテクスチャダイナミクスの モデル化が必要となる.

では,対象領域のダイナミクス(以下では単に領域ダイナミ クスと呼ぶ),特に形状の幾何学的変形は従来どのように扱わ れているだろうか.多くの場合,まず人物などの注目対象の形 状を輪郭によって記述し,その輪郭の時間変化を動的モデルと して表現する.例えば,文献[5]では,ジェスチャー中の人体 の輪郭について,輪郭線をサンプリングて得られた特徴点集合 の時系列に対して可変長マルコフモデルを学習している.また, 動的輪郭モデルなどを用いて物体の輪郭を推定,追跡するもの があり,医用画像における内臓の動きや表情の解析などに広く 用いられている[6].さらに,近年では対象の境界線をレベル セット関数によってimplicitに表現し,そのダイナミクスをモ デル化する手法も提案されている[7].

これらの領域ダイナミクスを表現する手法は,対象の追跡や 動作認識に対しては有効なモデルとなるが,一方で,テクス チャ変化を領域形状の変化と同時に扱うことはできない.

そこで本研究では、テクスチャダイナミクスと領域ダイナミ クスとを結びつけるひとつの枠組みとして、局所的なテクスチャ ダイナミクスに基づいて時空間領域を獲得し、一方で、領域ダ イナミクスに基づいて局所的なテクスチャダイナミクスの切り 替わりをモデル化するという新たな手法を提案する(図2). これにより、複数の異なる対象が含まれる動画像において、そ れぞれのテクスチャパターンの時間変化を精度良くモデル化で き、自然な風景映像の生成や動的背景モデルといった応用が可 能になると期待できる.





フレーム #12
フレーム #30
予測誤差(白いほど大)
図 4
各矩形小領域における線形システムの一期先予測誤差

### 2. 提案手法の概要

#### 2.1 2つのダイナミクスを結ぶ枠組み

まず, テクスチャダイナミクスを具体的に表現する動的モデル について述べる.入力された動画像 I(x, y, t) (x = 1, ..., W, y = 1, ..., H, t = 1, ..., T) に対して,空間方向に固定領域(窓)  $Z^{(r)} \subseteq [1, W] \times [1, H]$ を設定する(画像の全領域でもよい). この固定領域内の各時刻 t における輝度値を適当なスキャン方法でベクトル化することで,輝度値ベクトル $y_t^{(r)}$ を得る.すると,ベクトル系列  $y_1^{(r)}, ..., y_T^{(r)}$ から,以下の線形システムを同定することができる[1].

$$\begin{aligned} x_t^{(r)} &= A^{(r)} x_{t-1}^{(r)} + w_t^{(r)} \\ y_t^{(r)} &= C^{(r)} x_t^{(r)} + v_t^{(r)} \end{aligned} \tag{1}$$

ここで  $x^{(r)}$ ,  $A^{(r)}$ ,  $C^{(r)}$  は,それぞれ状態ベクトル,遷移行列,観測行列と呼ばれる.適当な初期値 $x_0^{(r)}$ から,上式を再帰的に適用することで,固定領域 $Z^{(r)}$ 内における輝度値の時系列  $\{y_t\}$ が得られる.これが,領域 $Z^{(r)}$ における Dynamic Texture のモデルである [1], [4].なお,特に断らない限り本稿では $x_t = [y_{t-p}^{\top}, ..., y_{t-1}^{\top}]^{\top}$ , C = [O, ..., O|I]のようなp次自己回帰モデルを具体的に用いるものとする.

ここで問題となるのは,風にゆれる木の葉などの対象をモデ ル化しようとする場合,類似したテクスチャダイナミクスを持 つ領域自体が,図3のように,時間的に変化することである. その結果,ある固定領域内のみを観測していると,異なるダイ ナミクスをもつ時区間が現れることになる.

図 4 は,各小領域のテクスチャダイナミクスを,それぞれ単 ーの線形システムで実際にモデル化した場合の一期先予測の誤 差である (T = 500[frame],各小領域は $32 \times 32$ [pixel]を $8 \times 8$ にダウンサンプリングして $y_t$ としている).白いほど予測誤 差が大きいことを表すが,特に木の葉と壁面が入れ替わる部分



図 5 各小領域が対象領域に含まれるか否かに基づいてテクスチャダ イナミクスを切り替える

で誤差が大きくなることが分かる.このような固定的なテクス チャとの切り替わりのほか,水面などの異なるテクスチャダイ ナミクスを持った背景と切り替わる場合や,さらには静止画と してのテクスチャがほぼ同じでも,異なる動きをする領域が入 れ替わる場合(大木の枝ごとで木の葉の動き)などがありうる.

そこで本研究では,まず画像を矩形小領域に分割した上で, 以下の枠組みを用いることにする(図5).

(1)類似したテクスチャダイナミクスが現れる小領域の集合を「対象領域」と定義する(一般には複数の対象があるが, 単純化のために対象と非対象の2つの領域があるとする).

(2) 個々の固定小領域では,対象領域に含まれる時間範囲 と含まれない時間範囲でそれぞれのテクスチャダイナミクスを 表現するような,2つの線形システムを同定しておく.

(3) 対象領域自体の形状変化(領域ダイナミクス)を別途 モデル化する.その上で,テクスチャ生成を行う場合には,

(4) 領域ダイナミクスに基づいて,個々の固定小領域が対 象領域に含まれるか否かを判定し,それぞれのテクスチャダイ ナミクスを表現する線形システムを切り替える

線形システムを時間的に切り替える手法としては,Switching LDS (Linear Dynamical System)がよく知られている[8]. 複 数の線形システムを HMM で切り替えることで,単一の時系 列を生成できるモデルである.一方,ここで提案するのは,各 領域における線形システムの切り替わりを,領域ダイナミクス に基づいて制御するという,空間的に拡張されたモデルであり, さらに複数の時系列を並列に生成していく点で大きく異なる.

2.2 問題の難しさと本手法におけるアプローチ

前節で述べたような枠組みを用いて,自然な風景映像の生成や,対象検出のための背景変動の将来予測を行うには,与えられた学習用の動画像において以下の問題を解く必要がある.

a) 動画像の時空領域分割

前節(1)における対象領域の推定は,テクスチャダイナミ クスの類似性に基づいた時空間領域分割の問題となる.仮に, 得ようとする時空間領域を表現する線形システムが既知であれ ば,動画像内においてその線形システムで精度良くモデル化さ れる領域として分割できる.しかし,実際には線形システムの



図 6 時空間ブロックの統合による時空間領域の獲得

パラメタは未知であり,逆にそのパラメタを推定するには,あ らかじめ同一のシステムで表現できる時空間領域が与えられる 必要がある.したがってこれは卵と鶏の問題であり,何らかの クラスタリング処理を要する.

文献 [4] の 2 次元領域分割を拡張した手法として,動画像を あらかじめ小さな時空間ブロックに分割しておき,それらの間 の類似性を定義した上で,(x,y,t)の 3 次元空間における領域 境界をレベルセット法で推定することが考えられる.しかし, 短時間のデータから同定したような線形システムに対して,シ ステム間の類似性を測ることは困難である.そこで,ここでは 図 6 のように線形システムの類似した時空間ブロック同士をボ トムアップに統合していくという手法をとることにする.具体 的には,類似性の定義も含め 3.節にて述べる.

b) テクスチャダイナミクスのモデル化

ー般に線形システムの同定では,学習データとして非常に長 い観測系列が与えられることを仮定する.しかし,前節(2)で 述べたように,テクスチャダイナミクスを線形システムの同定 によって得る際には,短時間だけ対象領域に含まれる小領域も ある.パラメタ数に対して比較的少量の学習データから,その テクスチャダイナミクスの線形システムを同定しようとすると 過学習となり,少しでも学習時と異なる状態に遷移した場合, 急激に状態が発散するなどの問題が生じる.

そこで,本手法では遷移行列に固有値制約を設ける手法[9] を利用する.これにより,比較的短時間の学習データからも安 定なシステムを同定することが可能となる(なお,本稿では紙 面の都合から 5.節で簡単に述べるにとどめる).

c) 領域ダイナミクスのモデル化

各時刻における対象領域の形状は,a)の領域分割時には,固 定小領域の集合として得られる.したがって,前節(3)にお ける領域ダイナミクスは,この小領域集合自体の時間変化を線 形システム等で記述することでモデル化できると思われる.し かし,小領域を単位としてモデル化した場合,生成時に各小領 域が連結していることが保証されないなどの問題が生じる.

そこで,ここでは閉じた境界を持つ領域のみを対象とするものとし,境界輪郭線の時間変化を線形システムで表現することにする.ただし,輪郭線を制御点やサンプリング点の集合で明示的に記述する場合,フレーム間での点の対応問題が生じる.これを避けるため,1.節で述べた,レベルセットによるimplicitな境界表現[7]を利用する(4.節参照のこと).これにより,領域形状の滑らかな変化が生成可能になるという利点もある.

## 3. テクスチャダイナミクスによる時空間領域 獲得

#### 3.1 領域成長法による時空間領域獲得

動画像 I(x,y,t) (x = 1,..,W,y = 1,..,H,t = 1,..T) が入 力されたときに,時空間領域  $\mathcal{R}$  を獲得したい.ただし,こ の段階では矩形小領域を単位として対象領域を記述する.そ こで,画像を矩形小領域  $Z^{(r)}$  (r = 1,..,R) に分割し,それぞ れの小領域の時刻 t における矩形ブロックを  $z_t^{(r)}$  とすれば,  $\mathcal{R} \subseteq \{z_t^{(r)} | t = 1,..,T, r = 1,..,R\}$  となる.以下では,領域成長 法によってこの時空間領域  $\mathcal{R}$  を獲得するアルゴリズムを示す. Step 0  $\mathcal{R}$  を空集合とする.

Step 1 各小領域  $Z^{(r)}$ におけるテクスチャダイナミクスを線 形システム(式(1))でモデル化する(Tフレームで同定). Step 2 各小領域における線形システムのモデル化誤差(-期先予測誤差)を求め,特に誤差の小さな小領域の1つ $Z^{(r_s)}$ を seed として選択する. $\{z_t^{(r_s)}|t=1,..,T\}$ を  $\mathcal{R}$ へ追加する. Step 3 既に得た時空間領域の近傍にある小領域集合 $\mathcal{N}(\mathcal{R})$ の中から,  $\mathcal{R}$ との距離  $\delta(Z^{(r)},\mathcal{R})$ が最小である小領域 $Z^{(r*)}$ を見つける.

Step 4  $\delta(Z^{(r*)}, \mathcal{R}) \leq Th_1$ なら  $\{z_t^{(r*)} | t = 1, .., T\}$ を  $\mathcal{R}$ に 追加し Step 3 に戻る.  $Th_1$ を超えるなら Step 5 へ進む.

Step 5 領域 R に含まれない各小領域  $Z^{(r)}$ の矩形ブロック系列を,固定長の時空間ブロック  $Z_k^{(r)} = \{z_t^{(r)} | t = b_k, ..., e_k\} (k = 1, ..., K)$ へと分割する.

**Step 6** 各時区間 k = 1, ..., K において順に次の処理を行う. (6-1) 近傍  $\mathcal{N}(\mathcal{R})$  にある小領域内で, $\delta(Z_k^{(r)}, \mathcal{R})$  が最小である時空間ブロック  $Z_k^{(r*)}$ を見つける.

(6-2)  $\delta(Z_k^{(r*)}, \mathcal{R}) \leq Th_2$  なら  $\{z_t^{(r*)} | t = b_k, ..., e_k\}$  を  $\mathcal{R}$  に追加 し (6-1) に戻る.  $Th_2$  を超えれば区間 k での処理は終了する. Step 7 全区間での処理を終了すれば,時空間領域  $\mathcal{R}$  を返す.

ここで Step 5 から先は,全フレームを一つの線形システム で表現できないような小領域に対して,その一部の時区間のみ を時空間領域 R へ併合する処理である.なお,上記のアルゴリ ズムでは,それまでに得られている時空間領域と,小領域(も しくは時空間ブロック)との距離 δ の定義が当然ながら重要と なるが,これについては次で述べる.

#### 3.2 線形システム間の距離の定義

小領域 Z<sup>(r)</sup>(もしくは時空間ブロック Z<sup>(r)</sup>)と時空間領域 R との距離は,それぞれを表現する線形システムを同定してお き,それら線形システム間の距離とする.この線形システム間 の距離は,直感的には以下のようになる.まず,一方の領域に 含まれる矩形ブロック系列から推定したシステムパラメタを用 いて,他方の矩形ブロック系列を一期先予測していき,そのと きの誤差を出す.この誤差は,パラメタ推定と誤差計算の系列 を逆にした場合でも算出できるため,これら二つの誤差の平均 を線形システム間距離とする.

小領域 Z<sup>(r)</sup> と時空間領域 R との場合で具体的に述べる.まず,それぞれの領域に含まれる矩形ブロック系列から,輝度値



図 7 獲得した時空間領域 (水面の動画像)(左: seed,右:獲得領域)

の系列  $\{y_t^{(r)}\}$  および  $\{y_t^{(\mathcal{R})}\}$  が得られる.今,線形システムとして,p 次自己回帰モデルを仮定しているため, $y_t^{(r)}$  からの線形システムの同定は,過去p フレームから現在のフレームを予測する誤差が小さくなるように,以下のように遷移行列  $A^{(r)}$ を推定すればよい.

$$\arg\min_{A^{(r)}} \sum_{t} ||y_t - CA^{(r)}[y_{t-p}^{(r)^{\top}}, .., y_{t-1}^{(r)^{\top}}]^{\top}||^2$$
(2)

ただし領域によらず常に C = [O, ..., O|I] とする(2.1節).  $\{y_t^{(\mathcal{R})}\}$ から遷移行列を推定する際も同様であるが,異なる小領 域に含まれていた矩形プロックを並べたものであるため,時間 的に不連続な部分は考慮する必要がある.

さて,時空間領域の推定パラメタ $A^{(\mathcal{R})}$ を用いて,小領域 $Z^{(r)}$ の系列を予測したときの誤差は

$$E_{r}(A^{(\mathcal{R})}) = \sum_{t} ||y_{t} - CA^{(\mathcal{R})}[y_{t-p}^{(r)^{\top}}, .., y_{t-1}^{(r)^{\top}}]^{\top}||^{2}$$

であり,逆に小領域の推定パラメタを用いて時空間領域の系列 を予測した  $E_{\mathcal{R}}(A^{(r)})$  との平均をとって領域間の距離とする.

$$\delta(Z^{(r)},\mathcal{R}) = \{E_r(A^{(\mathcal{R})}) + E_{\mathcal{R}}(A^{(r)})\}/2$$

#### 3.3 獲得された時空間領域の例

2 種類の動画像 (水面,木の葉) において,類似したテクス チャダイナミクスを持つ時空間領域をそれぞれ獲得した.まず, 水面の動きを含む動画像において獲得した時空間領域を図7に 示す.なお,獲得した時空間領域について,その空間的な領域 は時刻によらず図7のまま不変(一定領域)であったため,1枚 の画像のみを示している.水面におけるテクスチャ変化は一様 であるため,図7で示された時空間領域はおおよそ水面が占め ている領域と一致していることが分かる.

一方,図8は木の葉の揺らぎを含む動画像において獲得した 時空間領域である(画像は384×320,矩形領域は32×32).図 より,対象領域が時間の経過に伴い変化していく様子が分かる.

4. 領域ダイナミクスの動的モデル

#### 4.1 領域境界のレベルセット表現

3. 節の時空間領域より,対象領域の時間変化が得られる.これを学習データとして,領域形状の時間変化をモデル化するために,2. 節で述べたように,まず小領域単位で記述してきた対象領域をレベルセット関数  $\phi: \mathbf{R}^2 \to \mathbf{R}$ による表現とする[10].このとき領域の境界は

$$\{(x,y) \in \mathbf{R}^2 | \phi(x,y) = 0\}$$
(3)



図 8 獲得した時空間領域(木の葉の動画像)(フレーム番号は0開始)

で与えられる. *φ* としては,しばしば利用される符号付距離関 数<sup>(注1)</sup>を用いるものとする.なお,レベルセット法[10]ではこ の*φ* を伝播させることで対象の領域分割などを行うが,本稿で はこのような伝播は扱わず,あくまでレベルセット関数による 境界の表現のみを利用するものとする(6.3 節も参照のこと).

さて,領域の時間変化を表現するために, $\phi$ にフレームを表 す添え字をつけ, $\phi_t$ とする.3.節の時空間領域において,各フ レームでの符号付距離関数を求めることで $\phi_1, ..., \phi_T$ が得られ る.したがって,単純にはこの関数を適当なサンプリングで離 散化した上で,テクスチャダイナミクスの動的モデルと同様に, 過去数フレームの $\phi_{t-q}, ..., \phi_{t-1}$ から $\phi_t$ を推定できるようなモ デル化を行えばよいと考えられる.しかし,(x, y)を実際にど の程度のグリッドでサンプリングするかにもよるが,離散化さ れた $\phi_t$ はかなり高次元となる.

そこで,ここでは文献 [7] で提案されている, $\phi_t$ の基底表現 を用いて(通常の画像に対する固有空間法と同様)低次元に圧 縮する.まず,レベルセット関数の集合 { $\phi_1,..,\phi_T$ } に対する 主成分分析を行い,基底集合 { $\psi_1,..,\psi_n$ } を用いて

$$\phi_t = \phi_{\text{mean}} + \sum_{i=1}^N \alpha_{it} \psi_i \tag{4}$$

で表現されるようにする.ただし, $\alpha_{it}$ は $\phi_t - \phi_{\text{mean}}$ と基底 $\psi_i$ との内積で計算される.すると,主成分 $\alpha_{it}$ をベクトル系列とした $\alpha_1, ..., \alpha_T$ (ただし $\alpha_t = [\alpha_{1t}, ..., \alpha_{Nt}]^{\top}$ )を線形システムで表現すればよいことになる.

4.2 領域ダイナミクスのモデル化

系列 { $\alpha_t$ } に対する線形システムとして,これまでと同様に 自己回帰モデルを用いるものとする.ただし,文献 [7] に従い, 以下のようなバイアス項 ( $\mu$ ) をいれたモデルとする<sup>( $\pm 2$ )</sup>.

$$\alpha_t = \mu + A_1^{(\alpha)} \alpha_{t-1} + \dots + A_q^{(\alpha)} \alpha_{t-q} + \eta_t \tag{5}$$

生成時には,式(5)を用いて $\alpha_t$ を推定しながら,式(4)によっ

て  $\phi_t$  を求め, さらに式 (3) によって境界を得ることで, 領域 形状の時間変化が生成可能となる.

時空間領域の 50 から 250 フレーム目までを用いて領域ダイ ナミクスのシステムを同定し,初期値(次数 q = 3 としたため 最初の 3 枚)を与え,再帰的に  $\alpha_t$ を生成して得た領域の変化 を図 9 の下段に示す( $\alpha_t$ の次元数は 32 次元, $\eta_t = 0$  とした). 上段は学習系列であり,いずれも元画像に領域の境界線を重畳 表示している.生成時の領域の変化は学習系列とかなり一致す ることが分かる.

なお,学習時にはレベルセット  $\phi_t$ の段階で時間方向に平滑 化をするとモデル化精度が上がるため,ここでは学習系列に対 し5フレーム窓の移動平均を適用した.

## 5. 領域ダイナミクスによるテクスチャダイナミ クスの制御

領域ダイナミクスに基づいて,テクスチャダイナミクスを表 現する線形システムを切り替えながらテクスチャ生成を行う方 法について述べる.ただし,それに先立って,あらかじめ対象 領域内および領域外におけるテクスチャダイナミクスを,線形 システムによって表現しておく必要がある.

5.1 固有値制約を用いたテクスチャダイナミクスの同定

3. 節で獲得した時空間領域によって,各小領域 $Z^{(r)}$ は,対象領域内と領域外にある時区間に分節化される.対象領域内にあるような時区間の集合を $\{[b_k^{(in)}, e_k^{(in)}] | k = 1, ..., K_{in} \}$ とすれば,式(2)において,これらの区間内の時刻におけるベクトル $y_t$ が,直前のpフレームを用いて推定できるように,遷移行列 $A^{(r,in)}$ を推定する.一方,対象領域外についても同様に遷移行列 $A^{(r,out)}$ を推定する.つまり,画像内における小領域の数をRとすれば,最大で $R \times 2$ の遷移行列を推定することになる.

このとき,小領域によっては,大半の時間は領域外であり, 短時間だけ領域内となるようなもの(もしくはその逆)がある. 少数の学習サンプルから単純にシステム同定を行うと,推定さ れる遷移行列 A<sup>(r,in)</sup>(もしくは A<sup>(r,out)</sup>)によっては,生成系 列が急激に発散することがしばしば起こる.これは,遷移行列 の固有値の絶対値が1を超えた場合に生じる.そこで,固有値 の絶対値が1を超えないような制約を設けることで,少数の学

<sup>(</sup>注1):  $\phi(x, y)$  の値が, (x, y) から境界までの距離であり,境界内では正, 境界外では負の符号が付く

<sup>(</sup>注2): 今回の実験ではバイアス項の有無が結果に大きくは影響しなかった.



図 9 学習に用いた系列(上段)と生成結果(下段)



図 10 固有値の制約を設ける場合と設けない場合の比較(図1の小領 域で得られた遷移行列 A<sup>(r,in)</sup>の固有値および生成画像)

習サンプル(短い系列長)から,安定なシステムを同定するこ とが可能となる.詳細は文献[9]にゆずり,結果だけを見てみ ると,制約を設ける場合(図10左)は設けない場合(同図右) と比べて,固有値が複素平面の単位円の中に収まっており,さ らに生成画像列の輝度値も発散せずに木の葉から壁面へと移り 変わっていることが分かる.

5.2 領域ダイナミクスによるテクスチャダイナミクスの切替 4.2節で得られた領域ダイナミクスを用いることで,いった ん初期値が与えられれば,各時刻における対象領域の形状を生 成していくことができる.このとき,個々の固定小領域が対象 領域に含まれるか否かの判定を各時刻tで行う必要がある.そ こで,それぞれの小領域 $Z^{(r)}$ において,推定されたレベルセッ ト関数 $\phi_t(x,y)$ が正もしくは0となる面積の割合を求め,これ が閾値(例えば1/2)を超えれば対象領域内とすることにする. 具体例として,図1の矩形小領域におけるテクスチャ生成時の 実際の判定結果を図11に示す.

ここで,領域ダイナミクスによるテクスチャダイナミクスの 制御は,Heaviside 関数

$$H(\phi) = \begin{cases} 1 & \text{if } \phi \ge 0\\ 0 & \text{else} \end{cases}$$
(6)

を用いれば,テクスチャダイナミクスの線形システムを切り替えるための変数 $\gamma$ を介して以下のようにまとめることができる(記号の混乱を避けるため点(x, y)の代わりに $\xi \in \mathbf{R}^2$ を用いる).



図 11 図 1 の矩形小領域が対象領域に含まれるか否かの判定結果

**Region dynamics** 

$$\begin{cases} \alpha_t = \mu + \sum_{j=1}^q A_j^{(\alpha)} \alpha_{t-j} + \eta_t \\ \phi_t = \phi_{\text{mean}} + [\psi_1, .., \psi_N] \alpha_t \end{cases}$$
(7)

Texture dynamics

$$\gamma_t^{(r)} = \begin{cases} in & \text{if } \int_{Z^{(r)}} H(\phi_t) d\xi / \int_{Z^{(r)}} d\xi > 0.5 \\ out & \text{else} & (r = 1, ..., R) \end{cases}$$
(8)

$$\begin{cases} x_t^{(r)} = A^{(r,\gamma_t^{(r)})} x_{t-1}^{(r)} + w_t^{(r)} \\ y_t^{(r)} = C x_t^{(r)} + v_t^{(r)} \qquad (r = 1, ..., R) \end{cases}$$
(9)

なお,以下の評価実験では $w_t = 0, v_t = 0$ とした.また,矩形 プロック $z_t^{(r)}$ における特徴量 $y_t^{(r)}$ として,生成されるテクスチャ の精度を上げるために,領域内もしくは領域外として判断された 系列に対してそれぞれ主成分分析を行い,次元圧縮したベクトル を用いた.このとき,それぞれの基底集合を $\{u_1^{(r,in)},...,u_n^{(r,in)}\}$ ,  $\{u_1^{(r,out)},...,u_n^{(r,out)}\}$ ,平均を $m^{(r,in)},m^{(r,out)}$ とすれば,生成 される矩形ブロックの輝度値は以下のようになる.

$$z_t^{(r)} = m^{(r,\gamma_t^{(r)})} + [u_1^{(r,\gamma_t^{(r)})}, ..., u_n^{(r,\gamma_t^{(r)})}]y_t^{(r)} \quad (r = 1, ..., R)$$

5.3 評価実験

矩形小領域はいずれも  $32 \times 32$ とし, テクスチャダイナミク スの線形システムにおいては,自己回帰の次数 p = 3,  $z_t$ ,  $y_t$ ,  $x_t$ の次元をそれぞれ 1024, 16, 48 とした.一方,領域ダイナミク スは 4.2 節のパラメタをそのまま用いた.評価用のデータとし ては,図 9 の木の葉がゆれ動く動画像を用いた.学習時には 4.2 節と同様 200 フレームを用いて各システムのパラメタを推 定し,生成時には(1)一期先予測(250 フレーム,前半 200 フ レームは closed test 用,後半 50 フレームは open test 用),お よび(2) 初期値からの再帰的生成(すなわち線形システムの零

表 1 予測誤差 (1pixel あたりの平均二乗誤差)

Model	Closed test	Open test	Generation
Without switching	141.4	160.6	709.9
With switching	138.1	162.5	520.3

入力応答)(100フレーム)を行い,それぞれ元の動画像をどれ だけ再現できるかを評価した.

図 12 の左列は、テクスチャダイナミクスの線形システムを切 り替えない場合、右列は切り替えた場合における、生成画像列 の例である、切り替えない場合は、次第に木の葉のテクスチャ がぼけて行くとともに、木の葉のゆれが収束しているが、領域 ダイナミクスに基づいて線形システムを切り替えを行った場合 は、左右に木の葉がゆれ動き続けていることが分かる.

定量的な評価を行うために,一期先予測の closed test, open test, および初期値から再帰的に生成した際の, 1pixel あたり の平均二乗誤差を表1に示す.一期先予測では有意な差は見 られないが,生成を行った場合では,領域ダイナミクスを考慮 して線形システムを切り替える提案手法の方が,誤差が大幅 に低いことが分かる.さらに詳細な比較を行うために,矩形小 領域を,線形システムの切り替えが伴うもの (subregions with switching dynamics) と, それ以外 (subregions with constant dynamics) に分け,それぞれの小領域集合ごとで誤差の比較を 行った結果を図 13 に示す.特に線形システムの切り替えが生 じる小領域で,誤差が大きく減少しており,線形システムを切 り替えることの効果が確認できる.これは,図12でも見られ るように,領域ダイナミクスを用いない場合,次第に各小領域 のテクスチャが一定の値に収束していくが,一方で領域ダイナ ミクスを用いる場合は,木の葉のゆれといったより大域的な変 化を表現できるため,特に壁面や木の葉が交互に現れる部分な どにおいては,より精度良くテクスチャを生成できたと考えら れる.

#### 6. 討 議

提案手法の位置づけや限界を,関連手法と比較して考察する.

#### 6.1 Active appearance model との比較

テクスチャと領域形状とを結合する表現手法として,Active Appearance Model (AAM)がある[11].さらに,線形システ ムでその時間変化を表現することで,頭部や表情の追跡などに 用いることができる[12].表情のように,テクスチャと領域形 状との間に強い共起性があるような状況では,先に両者の結合 を考えた上で,その時間変化をモデル化するような AAM ベー スの手法が望ましいといえる.しかし一般には,領域形状とし ての変化は小さくても,領域内部のテクスチャが大きく変化す る状況がしばしば生じる(水面のテクスチャなど).提案手法 はこの後者の状況をターゲットとしたものであり,領域形状は, テクスチャと直接結合されるのではなく,テクスチャの時間発 展の仕方(いわばテクスチャダイナミクスのモード)のみを制 御するモデルとなっている.これら手法のいずれを選択するか は,モデル化対象となるシーンや物体の運動特性に依存する.







図 13 生成時における 1pixel あたりの平均二乗誤差(ダイナミクスの 切り替わりが生じる小領域群と生じない小領域群に分けた比較)

#### 6.2 輝度値ベースの領域分割法との比較

実験で用いた動画像では,対象領域内と領域外でテクスチャ や輝度値に明らかな違いが見られ,輝度値ベースの領域分割手 法[13] も利用可能であると思われる.しかし,2.1節で述べた ように,同じ木の葉であっても木の大枝ごとに異なる方向や周 期で動くなど,同一テクスチャだが異なるダイナミクスを持つ 領域が存在する状況も多い.映像生成や映像解析などの応用を 考えた場合,これらを別領域として分割・モデル化することが 望ましいが,本手法の枠組みでは,3.2節の線形システム間の 距離を用いることで,異なる速度や運動方向を持つ同一テクス チャ領域を分離可能であることを予備実験で確認している.

#### 6.3 レベルセット法による領域分割

4.2節では,領域境界をレベルセット関数によって表現して いるため,この表現を活かして,3.節の領域分割をレベルセッ ト法によって解くことが考えられる[10].そのためには,類似 性を定義する指標値(signature,例えば輝度値など)を予め各 画素で決定しておく必要があるが,本稿のように局所的ダイナ ミクスの類似性を signature にするには,画素周りの固定領域 を比較的長いフレーム数観測することになる[4].しかし,2.2 節(a)で述べたように,もし領域形状自体が時間的に変化する 場合,同じ固定領域に異なるテクスチャダイナミクスが現れる ため,適切な signature が得られないという問題がある.

3. 節のような領域成長法を用いた理由は,この解決策として, 領域分割と類似性 (signature) 計算を同時に行うことにあった. レベルセット法を導入する際にも同様に,signature を動的に 更新する方法を組み合わる必要がある.しかし現時点での問題 はむしろ,領域分割がボトムアップな処理のみによるところに ある.そこで例えば,(1)現在得られている時空間領域に基づ いて線形システムのパラメタと各画素のsignatureを更新,(2) 得られた signature を元にレベルセット法により領域獲得,を 繰り返すといった反復手法が有効と考えられ,その際には3.節 で得られた時空間領域を初期値として利用できる.このような 領域分割法の拡張に関しては,今後の課題としたい.

### 6.4 映像解析への応用

本稿では提案モデルの表現能力を確認するために,映像生成 を主な目的とした.一方で,映像中のダイナミクスを,3.節で 獲得した領域に基づいて調べることで,挙動検知や行動認識と いった映像解析に応用することが考えられる.5.節では,生成 されるテクスチャの精度を上げるために,小領域ごとに異なる 線形システムを同定したが,映像解析を目的とする場合,生成 テクスチャの精度はそれほど問題とならないため,領域内の各 小領域で同一のシステムパラメタを共有するとみなせばよい (3.節参照のこと).領域分割の高精度化が課題となるため,前 節の手法も含め,領域分割手法の拡張を検討する必要がある.

## 7. おわりに

対象領域の形状変化を表現する「領域ダイナミクス」のモデ ルと,固定領域における2次元信号の変化を表現する「テクス チャダイナミクス」のモデルとを結びつけることで,動画像を 生成する枠組みを提案した.

現時点では生成誤差も大きく,鑑賞に堪える十分な映像が得 られたわけではない.しかしその一方で,大域的な空間のダイ ナミクスによって局所的なダイナミクスを切り替えながらテク スチャ生成を行うという新たな枠組みは,今後様々な展開の余 地があると考えられる.現在,領域獲得から生成という一方向 処理だけではなく,生成に基づく領域獲得を反復的に組み合わ せる領域分割・モデル同定手法を検討しており,ダイナミクス に基づく映像解析などの応用へもつなげていく予定である.さ らに,本稿で扱った生成手法を発展させ,風を感じさせること ができる,より自然な映像生成の実現を目指す.

謝辞: 本研究の一部は,科学研究費補助金 18049046 の補助 を受けて行った.

文

#### 献

- S. Soatto, G. Doretto and Y. N. Wu: "Dynamic textures", Proc. IEEE Int. Conference on Computer Vision, pp. 439– 446 (2001).
- [2] J. Zhong and S. Sclaroff: "Segmenting foreground objects from a dynamic textured background via a robust kalman filter", Proc. IEEE Int. Conference on Computer Vision, pp. 44–50 (2003).
- [3] L. Yuan, F. Wen, C. Liu and H.-U. Shum: "Synthesizing dynamic texture with closed-loop linear dynamic system", Proc. European Conference on Computer Vision, pp. 603– 616 (2004).
- [4] G. Doretto, D. Cremers, P. Favaro and S. Soatto: "Dynamic texture segmentation", Proc. Intl. Conf. on Computer Vision, pp. 1236–1242 (2003).
- [5] A. Galata, N. Johnson and D. Hogg: "Learning variablelength Markov models of behavior", Computer Vision and Image Understanding, 81, 3, pp. 398–413 (2001).
- [6] A. A. Amini, R. W. Curwen and J. C. Gore: "Snakes and splines for tracking non-rigid heart motion", Proc. European Conference on Computer Vision, pp. 251–261 (1996).
- [7] D. Cremers: "Dynamical statistical shape priors for level set-based tracking", IEEE Trans on Pattern Analysis and Machine Intelligence, 28, 8, pp. 1262–1273 (2006).
- [8] V. Pavlovic, J. M. Rehg and J. MacCormick: "Learning switching linear models of human motion", Proc. Neural Information Processing Systems (2000).
- [9] H. Kawashima and T. Matsuyama: "Hierarchical clustering of dynamical systems based on eigenvalue constraints", 3rd International Conference on Advances in Pattern Recognition (S. Singh et al. (Eds.): ICAPR 2005, LNCS 3686), pp. 229–238 (2005).
- [10] L. A. Vese and T. F. Chan: "A multiphase level set framework for image segmentation using the Mumford and Shah model", Int. Journal of Computer Vision, 50, 3, pp. 271–293 (2002).
- [11] T. F. Cootes, G. J. Edwards and C. J. Taylor: "Active appearance model", Proc. European Conference on Computer Vision, pp. 484–498 (1998).
- [12] F. Dornaika and F. Davoine: "Simultaneous facial action tracking and expression recognition using a particle filter", Proc. IEEE Int. Conference on Computer Vision, pp. 1733– 1738 (2005).
- [13] D. A. Langan, J. W. Modestino and J. Zhang: "Cluster validation for unsupervised stochastic model-based image segmentation", IEEE Trans. on Image Processing, 7, 2, pp. 180–195 (1998).