

線形システムに基づく動的背景のモデル化と移動対象検出への応用

三井 健[†] 波部 斉^{††} 松山 隆 司

概要 動画画像から移動対象物体を検出するためには、あらかじめ背景シーンの変動を学習しておく必要がある。本論文では、線形システムを用いて背景変動を線形な自己回帰モデルで表現する。精度の高いモデル化を行うため、提案手法では入力画像に対して平滑化を行い、解像度を粗くすることで予測誤差を低減させる。実験により、平滑化を行うことで予測誤差が低減され、高精度の対象検出が実現できることが示された。また、変動が小さく規則的な場合と、変動が大きく不規則な場合を比較すると、モデル化誤差を十分小さくするには後者のほうが解像度をより粗くする必要があることがわかった。

Dynamic Background Modeling using Linear Dynamical System

KEN MITSUI, [†]HITOSHI HABE ^{††} and TAKASHI MATSUYAMA[†]

Abstract In order to detect target objects in dynamic scenes, we need to model a non-stationary backgrounds. In this paper, we introduce the linear dynamical system (LDS) as a model to express the non-stationary background. LDS consists of a linear equation of a state vector, which represents a continuous state transition. It is quite important how to obtain the state vector from observed images, because the inference accuracy of the background variations depends on the state vector. To obtain the suitable state vector, we have to reduce the inference error. Therefore, we smooth observed images and get the state vector from smoothed images. This is because such a state vector reflects only macro motion of the background scene, and would provide us a more accurate inference. Experimental results demonstrate the effectiveness of our method.

1. はじめに

1.1 研究の背景

実世界のシーンを撮影した動画画像からの人・車両・障害物などの検出は、視覚情報処理において基本的かつ必要不可欠な処理である。たとえば、画像中の認識対象の位置を求めたり、認識対象から何らかの情報を得ようとしたりする場合、対象が占める領域を把握し、非対象領域を除外したうえで認識をおこなうことが必要となる。また、実社会でのセキュリティ対策における必要性の高まりもあって、対象検出に関する研究は盛んに行われてきた。

他方、計算機の処理能力の向上に伴い、大量のデータに対する統計的な処理が現実的なものとなっている。対象検出においても、大量の動画画像に対する処理が可能となり、それらに対して統計的な解析手法を用いるものが主流となっている。

1.2 先行研究

対象検出は、画像を検出対象とそれ以外(背景)との二つの領域に分別する操作と考えられ、その方法には対象モデルに基づくものと背景モデルに基づくものがある。対象モデルに基づく対象検出は対象が既知でなければならぬという制約があるが、背景モデルに基づく対象検出は背景についてのみ学習するので、対象は未知であって構わない。そのため、実世界における対象検出の中で、対象に関して未知である場合にはこの方法が有効である。このように、背景モデルに基づいた対象検出は、対象検出を考える上で最も基本的かつ有効なアプローチである。実際、実用化されているシステムでも、背景モデルに基づいた処理を行ったあとで可能であれば対象モデルに基づく検出を行う、という設計のものが多く、したがって本論文では背景モデルに基づいた対象検出を取り上げることとする。

背景モデルに基づく対象検出では、背景に関する情報を学習する必要がある。通常、実世界では背景の状態は不変でないため、単純な背景差分に基づく対象検出は有効ではない。そこで、実世界における対象検出を実現するには、背景の変動を適切にモデル化する必要がある。背景変動のモデル化の方法としては、これまで様々なものが提案されてきた。それらは以下の3

[†] 京都大学大学院情報学研究所
Graduate School of Informatics, Kyoto University

^{††} 京都大学大学院工学研究所
Graduate School of Engineering, Kyoto University

つに大別される¹⁾。

(1) 背景変動に対して不変な特徴を利用する方法。例として、画像をベクトル表記し、正規化することでテクスチャパターンを記述するものが挙げられる²⁾。(2) 与えられた背景画像系列より計算される統計量を用いる方法。統計量として画素ごとの平均値、中央値などを用いるもの³⁾、画素値に対する背景としての確率を正規分布で表現するもの⁴⁾などがある。(3) 背景変動の時間的、空間的な相関関係を注目する方法。前者の例として、各ブロックにおける入力画像と基準画像との相関値が時間経過に伴ってどう変化するかを解析し、背景変動パターンに応じた背景差分処理を行う方法が提案されている⁵⁾。また後者の例として、隣接する2つのブロックに強い相関関係があると仮定し、着目しているブロックの背景らしさを隣接ブロックから推定するという方法が挙げられる⁶⁾。時空間双方の相関関係を用いたものとしては Toyama らによる Wallflower⁷⁾ が知られており、そこでは画素の時間変化を線形フィルタで表現し、空間的な隣接関係を考慮して性能向上を図っている。(1)は時間変化に対して不変な特徴を単独の画像から抽出するものであるが、現実的には完全に不変な特徴を抽出することは不可能であるので、その時間変化を(2)のような統計モデルで記述するものが用いられている。さらに、単純な統計量ではなく、時間的あるいは空間的な変化・相関関係を利用したものが(3)であるといえ、順をおってモデルの複雑さは増している。

1.3 線形システムを用いた対象検出

このような中で、本論文では線形システムを用いて背景画像の時間的変化を記述し、背景変動のモデル化を行う。ここでいう線形システムとは、2章で述べるとおり、内部状態を示す状態ベクトルの変化が線形な方程式で記述できるシステムのことを指し、時刻 t での状態ベクトルを、時刻 $t-1$ から $t-R$ までの状態ベクトルから推定するというものである。線形システムを用いた研究例として、川嶋らは、唇映像の区間への分節化を行う際、各区間は線形システムで表現できるという仮定のもと、分節化された区間のタイミング構造を用いて人間の表情の記述・生成・認識を行っている⁸⁾。

背景変動の要因として、実環境では(1)照明条件の変化、(2)背景に存在する物体の移動・出現、(3)木の葉の揺らぎ、周期的・規則的とみなせる変化、など様々なものが挙げられるが、本論文ではこのうち(3)に着目して背景変動をモデル化する。(3)の変動は、過去から連続して発生するものであり、それを表現するものとして線形システムは妥当なものであるといえる。さらに線形システムでは、演算がすべて線形な行列演算で記述できるため単純かつ高速であり、実時間処理が必要とされるシステムへの適用にも有利であるといえる。このような観点から、本論文では動的背景の変

動を記述する方法として線形システムに着目する。

本論文でとりあげる動的背景のモデル化に近いものとしては、Doretto らによる Dynamic Textures⁹⁾があげられる。そこでは、時間経過とともに変動するテクスチャを線形システムを用いて記述し、テクスチャの合成・識別・圧縮などを行っている。しかし次章以降で述べるとおり、線形システムで表現できる状態ベクトルの変動は限られており、状態ベクトルの選択によっては予測精度が低下し、対象検出精度の悪化につながる。したがって線形システムを用いる際には、観測される画像データから状態ベクトルへの変換をどのように行うかが非常に重要な問題となる。これに対し本論文では、以下の2つのアイデアによって問題の解決を図る。

- (1) 観測画像に対して平滑化による解像度変換を施すことによって、線形システムで精度よくモデル化可能な状態ベクトルを得る。
- (2) 線形システムのモデル化誤差だけではなく、状態ベクトルの変動を解析することで、対象検出に最適な状態ベクトルを選択する。

細かい解像度の観測画像をそのまま状態ベクトルとしたときには、それぞれの要素の変動が複雑になって線形システムで表現したときの予測誤差が大きくなるが、それを平滑化して得た状態ベクトルでは大域的な変動のみが表現されているので、線形システムでの予測誤差の低減が期待できる。しかし、極端に平滑化した場合には予測誤差が小さくなくても検出対象と背景の識別が不可能になる可能性があるので、(2)のように状態ベクトルの変動を解析して最適な解像度を求める。

以下、2章において本研究で取り上げる線形システムとそれを用いた対象検出処理について述べる。3章で画像の解像度に着目することで対象検出に適切な状態ベクトルを推定する方法を提案する。4章では提案手法の実験結果を示しその有効性を明らかにする。最後に5章で本研究の成果をまとめ、今後の課題を論じる。

2. 線形システムを用いた背景変動のモデル化

本章では、線形システムを用いた背景変動のモデル化の方法について述べる。

2.1 線形システム

本節では、線形システムの具体的な定義について述べる。ここで、まず以下の用語と記法を定義する。

観測ベクトル： 観測データから直接得られるベクトルが観測ベクトルであり、時刻 t における観測ベクトルを y_t と表す。

状態ベクトル： あるシステムの状態を示すベクトルが状態ベクトルであり、観測ベクトルを変換することで得られる。時刻 t における状態ベクトルを x_t と表す。

線形システムとは、状態ベクトルの時間的変化が以下の線形な方程式により記述できるとするものである。

$$\mathbf{x}_t = \mathbf{F}\mathbf{x}_{t-1} + \mathbf{w}_t \quad (1)$$

$$\mathbf{y}_t = \mathbf{G}(\mathbf{x}_t) + \mathbf{v}_t \quad (2)$$

式(1)を状態方程式、式(2)を観測方程式と呼ぶ。状態方程式は、時刻 t での状態ベクトルを時刻 $t-1$ での状態ベクトルから決定するため、1次の自己回帰モデルであるといえる。また、 \mathbf{F} は状態遷移行列、 \mathbf{G} は観測空間と状態空間を結びつける関数であり、 \mathbf{w}_t はプロセスノイズ、 \mathbf{v}_t は観測ノイズを表す。本研究では、 \mathbf{w}_t 、 \mathbf{v}_t を無視する。

\mathbf{F} を、状態系列 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T$ から決定することを考える。 $\mathbf{X}_0 = [\mathbf{x}_1, \dots, \mathbf{x}_{T-1}]$ 、 $\mathbf{X}_1 = [\mathbf{x}_2, \dots, \mathbf{x}_T]$ とおくと、 \mathbf{F} の決定は、各時刻における二乗予測誤差を最小にする問題と考えることができる。すなわち、

$$\mathbf{F}^* = \arg \min \|\mathbf{F}\mathbf{X}_0 - \mathbf{X}_1\|^2 \quad (3)$$

と表せる。ここで \mathbf{F}^* は \mathbf{F} の近似解である。これを解いて、

$$\mathbf{F}^* = \mathbf{X}_1\mathbf{X}_0^T(\mathbf{X}_0\mathbf{X}_0^T)^{-1} = \mathbf{X}_1\mathbf{X}_0^+ \quad (4)$$

となる。 \mathbf{X}_0^+ は \mathbf{X}_0 の一般化逆行列である。

ここで、本論文では状態方程式(1)を以下のように定義しなおす。

$$\mathbf{x}_t = \mathbf{F}[\mathbf{x}_{t-1}, \mathbf{x}_{t-2}, \dots, \mathbf{x}_{t-R}]^T \quad (5)$$

式(5)は自己回帰モデルの次数を R に増やしたものである。次数を増やすことにより、直前の状態変化の度合(状態変化の「速度」など)を加味できるので、より正確に状態変化を記述できると考えられる。

2.2 背景変動のモデル化と対象検出

本節では、背景変動のモデル化、および対象検出の方法について述べる。

画像内の背景変動において、ある程度の大きさを持った物体の変動を記述するためには画素単位の変動に注目しても有益な情報は得られにくく、ある程度大きい範囲に注目するほうが物体の変動を捉えやすい。一方、画像全体をひとまとめにして表現すると情報量が膨大になってしまう。また、物体の変動を記述するにはその周辺の局所的な範囲に注目すれば十分であり、物体と離れた箇所の変動は相関が低いと考えられる。したがって本論文では、画像をブロックに分割し、ブロック内の画像パターンを特徴量としてブロック単位で背景変動をモデル化するという方法をとる。

各ブロックにおける背景変動をモデル化し、対象検出を行うまでの流れを以下に示す(図1, 2)。

- (1) 観測画像 $I_t(t=1, \dots, T)$ を $N \times N$ 画素のブロックに分割し、各ブロックを $I_t^{(k)}$ とする。添字 k はブロック番号を表す。
- (2) 画像ブロック $I_t^{(k)}$ を N^2 次元の列ベクトルで表記したものを観測ベクトル $\mathbf{y}_t^{(k)}(t=1, \dots, T)$ とし、これを n 次元の状態ベクトル $\mathbf{x}_t^{(k)}(t=1, \dots, T)$ に変換する ($n \leq N^2$)。状態ベクトル

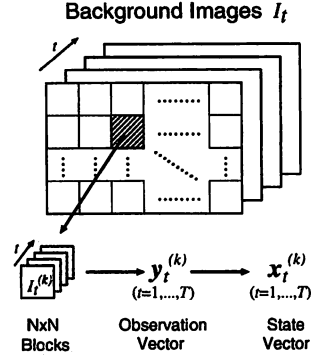


図1 画像の分割と状態ベクトルへの変換。

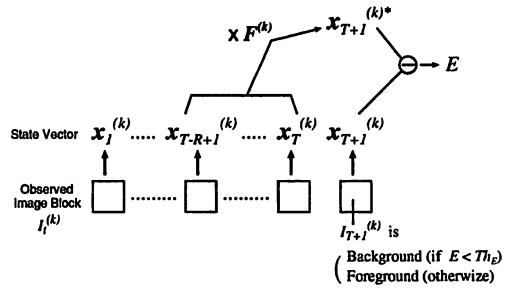


図2 背景か対象かの識別。

ルへの変換法については3章で述べる。

- (3) 得られた状態ベクトル $\mathbf{x}_1^{(k)}, \mathbf{x}_2^{(k)}, \dots, \mathbf{x}_T^{(k)}$ から遷移行列 $\mathbf{F}^{(k)}$ を算出する(2.1節参照)。
- (4) 得られた状態ベクトル $\mathbf{x}_t^{(k)}(t=1, \dots, T)$ が、状態方程式(5)をほぼ満たすと仮定する。このとき、 $t=T+1$ における状態ベクトルが次式から予測できる。

$$\mathbf{x}_{T+1}^{(k)*} = \mathbf{F}^{(k)}[\mathbf{x}_T^{(k)}, \mathbf{x}_{T-1}^{(k)}, \dots, \mathbf{x}_{T-R+1}^{(k)}]^T \quad (6)$$

ここで、 $\mathbf{x}_{T+1}^{(k)*}$ は $t=T+1$ における予測される状態ベクトルを表す。

- (5) $\mathbf{x}_{T+1}^{(k)*}$ と、実際に得られる状態ベクトル $\mathbf{x}_{T+1}^{(k)}$ との差を、ブロック $I_{T+1}^{(k)}$ が背景か対象かを識別する尺度に用いる。すなわち、 $\mathbf{x}_{T+1}^{(k)*}$ と $\mathbf{x}_{T+1}^{(k)}$ との誤差を E としたとき (E の定義については3.2.1節参照)。

- $E < Th_E$ ならばブロック $I_{T+1}^{(k)}$ は背景
- $E \geq Th_E$ ならばブロック $I_{T+1}^{(k)}$ は検出対象と判断する (Th_E は閾値)。

以上の方法により、ブロック単位でそのブロックが背景であるか対象であるかを識別する。

なお、Step(2)において観測ベクトル $\mathbf{y}_t^{(k)}$ から状態ベクトル $\mathbf{x}_t^{(k)}$ を生成する必要があり、 $\mathbf{x}_t^{(k)}$ から $\mathbf{y}_t^{(k)}$

への変換は必要でないため、本論文では観測方程式(2)を以下のように改める。

$$\mathbf{x}_t^{(k)} = H(\mathbf{y}_t^{(k)}) \quad (7)$$

関数 H により $\mathbf{y}_t^{(k)}$ から $\mathbf{x}_t^{(k)}$ への変換が可能となる。関数 H の内容、すなわち Step(2)における画像ブロックから状態ベクトルを生成する過程は、線形システムを用いるにあたって非常に重要なものであり、どのような生成法を用いるかについて十分考察する必要がある。

3. 画像の平滑化による最適な状態ベクトルの選択

本章では、背景変動のモデル化において必要となる、画像ブロックから状態ベクトルを生成する方法について述べる。

3.1 平滑化画像からの状態ベクトルの生成

状態ベクトル $\mathbf{x}_t^{(k)}$ を生成する方法についてはさまざまなものが考えられるが、線形システムを用いるうえで適切なものを選択しなければならない。そのため、状態ベクトルの生成法として、状態ベクトルが線形システムを用いて精度よくモデル化可能となるものを選択することが少なくとも必要となる。

状態ベクトルとして最も単純なものとしては、画像ブロックをそのままベクトル表記したものが考えられるが、この場合状態ベクトルの各要素はもとの画像の画素を表しており、それぞれの要素の変動が非常に複雑なものとなる。そのため状態ベクトルの変動を線形システムで表現したときの高い予測精度は得られにくい。

そこで、画像の解像度を粗くしてから状態ベクトルを獲得すれば得られた状態ベクトルは背景の大域的な変動のみを表現しているため、線形システムに適用したときに誤差の小さい予測が行えると考えられる。そこで本論文では、状態ベクトルを生成する過程において画像の解像度を粗くするという操作を行うことで、線形システムで精度よくモデル化可能な状態ベクトルを得ることを試みる。

以下、画像ブロック $I_t^{(k)} (t = 1, \dots, T)$ から状態ベクトル $\mathbf{x}_t^{(k)}$ を生成するアルゴリズムを示す(図3)。

- (1) 画像ブロック $I_t^{(k)} (N \times N$ 画素)内に、 $M \times M$ 画素の小ブロックを考える。小ブロックの始点を画像ブロックの左上端(座標を $(0,0)$ とする)とし、小ブロック内の画素値の和を x'_1 とする。なお、 M をどの程度にすればよいかの議論は次節で述べる。
- (2) 小ブロックの始点を $(0,0), (0,1), \dots, (1,0), (1,1), \dots, (N-M, N-M)$ と移動させていき、それぞれの始点の場合における小ブロック内の画素値の和を x'_1, x'_2, \dots, x'_Z とする(ここで、 $Z = (N-M+1)^2$)。

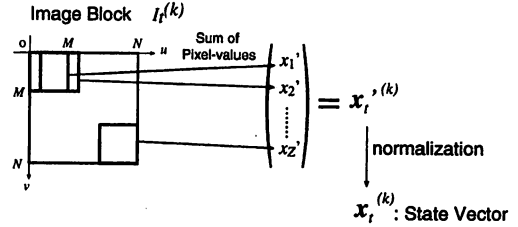


図3 解像度変換による状態ベクトルの生成。

- (3) ベクトル $\mathbf{x}_t^{(k)} = [x'_1, x'_2, \dots, x'_Z]^T$ を、時刻 t における状態ベクトルとする。状態ベクトルの各要素が小ブロック内の画素値の和であるので、以上の操作は画像の解像度を小ブロックサイズのレベルまで粗くすることと同等である。
- (4) 状態ベクトル $\mathbf{x}_t^{(k)}$ を以下のように正規化する。

$$\mathbf{x}_t^{(k)} = \frac{\mathbf{x}_t^{(k)}}{\|\mathbf{x}_t^{(k)}\|} \quad (8)$$

得られた $\mathbf{x}_t^{(k)}$ を、改めて状態ベクトルとする。正規化することにより、一様な画素値の変化の影響を受けなくなる。照明条件の変化が画素値の一様な変化を引き起こすと近似できるとすると、正規化により照明条件の変化の影響を削減することができる。

以上の方法により、線形システムで精度よくモデル化が可能な状態ベクトルを得ることができると考えられる。

3.2 対象検出に適切な解像度の選択

前節の方法により得られる状態ベクトルは、小ブロックのサイズ M 、すなわち画像の解像度に依存する。 M のとりうる値の範囲は $1 \leq M \leq N$ であり、解像度に応じて得られる状態ベクトルは異なるため、どの状態ベクトルを用いるかによって対象検出の精度はそれぞれ異なる。そのため各ブロックにおいて対象検出を行う際に最も適切な解像度が存在する。したがって、各ブロックにおいて最適な解像度を推定できれば、その解像度において得られる状態ベクトルを用いることで、高精度の対象検出が行えると期待できる。本節では、各ブロックで最適な解像度を推定する方法を提案する。

3.2.1 状態ベクトルの予測誤差

対象検出を行う上で最適な解像度を選択する際に最も重要となるのが、その解像度において得られる状態ベクトルの変動が、線形システムで精度よく表現できることである。

画像の解像度が細かい場合、その画像から得られる状態ベクトルは各要素の変動が複雑であり、線形システムで表現したときに高い予測精度は得られにくい。一方画像の解像度が粗い場合は、得られる状態ベクトルは大域的な変動のみを表現しており、状態ベクトル

を線形システムに適用したときその変動が精度よく表現できると予想できる。

本論文では、状態ベクトルの変動が線形システムで精度よく表現できるかどうかの尺度として、次式で表される E の値を用いる。

$$E(\mathbf{x}^{(k,M)}) = 1 - \frac{1}{T-1} \sum_{t=2}^T \text{Sim}(\mathbf{F}^{(k,M)} \mathbf{x}_{t-1}^{(k,M)}, \mathbf{x}_t^{(k,M)}) \quad (9)$$

ただし、 $\text{Sim}(\mathbf{a}, \mathbf{b})$ はベクトル \mathbf{a}, \mathbf{b} 間の類似度を表す関数であり、

$$\text{Sim}(\mathbf{a}, \mathbf{b}) = \frac{\langle \mathbf{a}, \mathbf{b} \rangle}{\|\mathbf{a}\| \|\mathbf{b}\|} \quad (10)$$

と定義する ($\langle \cdot \rangle$ は内積)。この値が大きいほどベクトル \mathbf{a} と \mathbf{b} の類似度が高い。

すなわち E は予測した状態ベクトル $\mathbf{x}_t^{(k,M)*}$ と実際の状態ベクトル $\mathbf{x}_t^{(k,M)}$ との予測誤差を意味する。得られた状態ベクトルの E が小さいほど予測誤差は小さく、その状態ベクトルの変動が線形システムで精度よく表現できると考えられる。

各ブロックにおいて複数の解像度における E の値を求め、 E の値を最適な解像度推定のための評価値の1つとする。

3.2.2 状態ベクトルの変動

上で述べたように、 E が小さいほどその状態ベクトルの変動が線形システムで精度よく表現できると考えられる。しかし、対象検出に最適な解像度を決定するための評価値として E だけでは不十分である。

E が小さければ状態ベクトルの変動の予測精度は高いといえるが、 E が小さい場合として、単に画像が極端に平滑化されているために状態ベクトルの変動がほとんど見られずほぼ不変であり、そのために予測精度が高いという可能性もありうる。

仮に、ある画像ブロックに対し異なる2つの解像度において状態ベクトルを得たときに E が同程度であった場合、状態ベクトルの変動が小さいほうは単に極端に平滑化されているために E が小さいと考えられる。他方、変動が大きい状態ベクトルは細かい背景変動まで表現しているといえるので、 E が同程度の場合は変動の大きい状態ベクトルを用いるのが適切と考えられる。

また、極端に平滑化したために状態ベクトルの変動が小さい場合には、対象が現れた場合に背景との識別が困難になり、対象検出の精度が低下する可能性がある。そのような観点からも、状態ベクトルの変動に注目することは重要といえる。

状態ベクトルの変動を評価するものとして、次式で表される変動値 V を用いる。

$$V(\mathbf{x}^{(k,M)}) = \frac{V_0(\mathbf{x}^{(k,M_{min})})}{V_0(\mathbf{x}^{(k,M)})} \quad (11)$$

ここで M_{min} は、 M の値を変化させていくときの M

の最小値であり、 V_0 は

$$V_0(\mathbf{x}^{(k,M)}) = \frac{1}{T-1} \sum_{t=2}^T \text{Sim}(\mathbf{x}_{t-1}^{(k,M)}, \mathbf{x}_t^{(k,M)}) \quad (12)$$

で定義される。 $\text{Sim}(\cdot)$ は $\mathbf{x}_{t-1}^{(k,M)}$ と $\mathbf{x}_t^{(k,M)}$ の類似度を表し (式 (10) 参照)、この値が小さいほど時刻 $t-1$ から時刻 t での状態ベクトルの移動が大きい。そのため V_0 が小さいほど状態ベクトルの変動が大きいものとみなせる。

したがって V は、同一ブロックにおける、解像度 M での状態ベクトルの変動と解像度 M_{min} での状態ベクトルの変動との比を示し、解像度を変化させて V の値がどう変化していくかを調べることにより、状態ベクトルの変動が解像度 M_{min} の場合に比べてどれだけ小さくなっていくかを解析することができる。この V の値を、最適な解像度推定のための2つめの評価値とする。

3.2.3 解像度選択アルゴリズム

以上のように、各ブロックにおいて対象検出に最適な解像度を推定するための評価値として、各解像度において得られた状態ベクトルの予測誤差 E 、および状態ベクトルの変動値 V を用いる。

対象検出を行うための最適な解像度を選択するにあたって、状態ベクトルを線形システムに適用して予測を行うことからその予測誤差 E は小さい必要があり、 E が大きいと線形システムは精度よく機能しない。また、 E が小さくても変動値 V が小さければ単に画像が極端に平滑化されているだけの可能性があるため、 E が小さいものの中で V は大きいほうが好ましい。以上をふまえると、対象検出を行うための最適な解像度を選択するアルゴリズムは次のようになる。

- (1) 各ブロックにおいて、複数の解像度に対してそれぞれの場合に得られる状態ベクトルの予測誤差 E 、変動値 V を求める。
- (2) 解像度の中で、 $E < Th_E$ を満たすものを選択する (Th_E は閾値)。満たすものが無い場合は、 E が最小である解像度を選択する。
- (3) Step(2) で選択された解像度がただ1つなら、その解像度を採用する。2つ以上ある場合は、 V が最大であるものを採用する。

以上の方法により、各ブロックにおいて対象検出に最適な解像度が推定できる。この解像度において得られた状態ベクトルを線形システムに適用することで高精度の対象検出を行えると期待できる。

4. 実 験

本章では、3章で述べた手法により状態ベクトルを選択したときの対象検出の精度を実験により評価する。

4.1 画像パターンと予測誤差の関係

入力画像 (図4) の各ブロックの予測誤差を濃淡値で表したものを図5に示す。黒に近いほど誤差が小さく、

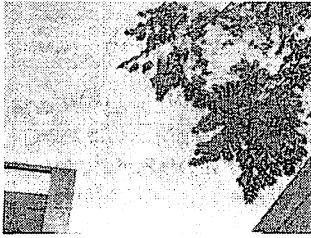


図4 入力画像.

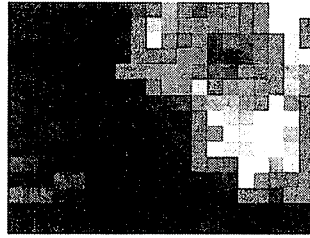


図5 予測誤差の濃淡表示.

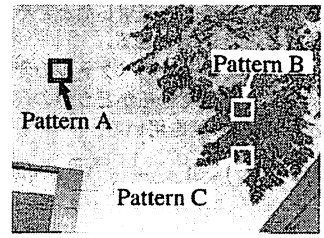


図6 画像ブロックのパターン.

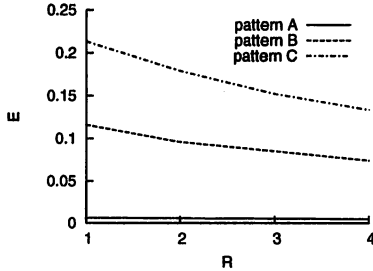


図7 次数と予測誤差の関係。
ただし E の値を 10^4 倍で示してある。

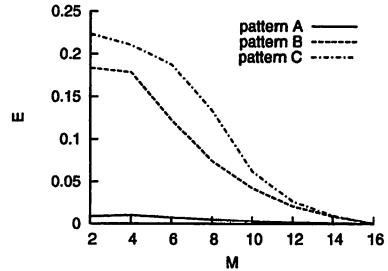


図8 解像度と予測誤差の関係。
ただし E の値を 10^4 倍で示してある。

白に近いほど誤差が大きいことを意味する。この結果より、空や建物を表すブロックでは誤差が小さく、木の葉の部分や、木の葉と空の境界部分では誤差が大きいことが読みとれる。したがって、本論文では画像ブロックを、その表すシーンによって

- パターン A 静止シーンを示すブロック (空や建物)
- パターン B 木の葉の揺らぎのシーンを示すブロック
- パターン C 静止部分と木の葉の境界に位置するブロック

の3つに大別し (図6)、各パターンごとのブロックの特徴を分析する。

4.2 状態ベクトルの選択とその精度評価

本論文において、状態ベクトルを決定するパラメータは (1) 自己回帰モデルの次数 R 、(2) 画像の解像度 (小ブロックサイズ M)、の2つである。本節では、これらのパラメータを変化させることにより、得られる状態ベクトルが線形システムで精度よくモデル化できるものであるかどうかを予測誤差 E や変動度 V を尺度として評価する。

4.2.1 自己回帰モデルの次数と予測誤差の関係

状態方程式 (式 (5)) における自己回帰モデルの次数 R と、そのときに得られる状態ベクトルの予測誤差 E の関係を調べた。解像度は $M = 8$ として統一した。各パターンのブロックにおいて $R = 1, 2, 3, 4$ としたときの R と E の関係を表したグラフを図7に示す。横軸は R 、縦軸は E である。

パターン A のブロックでは、 R に関わらず E は極めて小さい値のままほぼ一定であったが、これは静止

シーンでは変動がほとんど無いために予測が容易であるためと考えられる。パターン B、C のブロックでは、ほぼ全てのブロックにおいて R を大きくするに従って E は小さくなった。この結果より、自己回帰モデルの次数を増やすことで予測誤差を低減できることが確認できる。理由としては、直前の状態変化の割合を加味することで予測精度が向上するためと考えられる。

4.2.2 画像の解像度と予測誤差の関係

画像の解像度 M と、その解像度において得られる状態ベクトルの予測誤差 E の関係を調べた。自己回帰モデルの次数は $R = 4$ とした (以下の実験でも同様)。各パターンのブロックにおいて、 $M = 2, 4, \dots, 14, 16$ としたときの M と E の関係を表したグラフを図8に示す。

パターン A は静止シーンであるため、 E は極めて小さい値のままほぼ一定であった。一方パターン B、C のブロックでは、解像度を粗くするに従って E が減少するものがほぼ全てであった。理由としては、平滑化した場合、得られる状態ベクトルでは大域的な変動のみが表現されるので予測誤差は小さくなるためであると考えられる。また、パターン C のほうが B に比べて E は大きいという結果が得られたが、パターン C のブロックでは木の葉が入り出すため変動が不規則であり、予測が難しいためと考えられる。以上の結果より、画像の解像度を粗くすることにより線形システムのモデル化誤差が小さい状態ベクトルが得られることが確認できる。

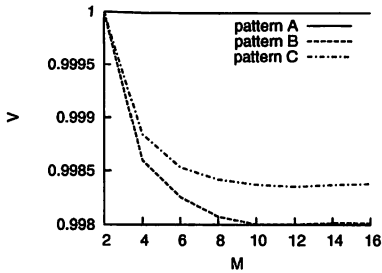


図9 解像度と変動値の関係。

4.2.3 画像の解像度と状態ベクトル変動の関係

画像の解像度 M と、その解像度において得られる状態ベクトルの変動値 V の関係を調べた。各パターンのブロックにおいて $M = 2, 4, \dots, 14, 16$ としたときの M と V の関係を表したグラフを図9に示す。

パターンAのブロックでは V はほぼ一定であったが、静止シーンではそもそも状態ベクトルの変動が極めて小さいためと考えられる。パターンB、Cのブロックでは解像度を粗くするに従い V が小さくなり、解像度が十分粗くなると V はほぼ一定となった。平滑化するに従い、得られる状態ベクトルでは細かい変動は徐々に表現されなくなっていく、大域的な変動のみが表現されるためと考えられる。また、パターンCのほうがBに比べて平滑化に伴う V の減少率は小さかったが、この理由としてパターンCのブロックは静止部分を含んでおり、その部分の変動は解像度によらず小さいためと考えられる。

4.2.4 まとめ

自己回帰モデルの次数 R については、 R を大きくすることで状態ベクトルの予測誤差を低減できる。

画像の解像度に関しては、解像度を粗くすることにより得られる状態ベクトルの予測誤差 E は小さくなり、表現できる変動は大域的なものに限られるために状態ベクトルの変動値 V は小さくなる。

4.3 線形システムを用いた対象検出の精度評価

画像の解像度を粗くすることで得られる状態ベクトルの予測誤差は小さくなるが、対象検出における背景と対象との識別能力が低下する可能性がある。これについて検証するために、ここでは解像度を変化させることで対象検出の精度がどのようになるかを調べた。

なお、本実験では、対象物体を含む画像を用いず、背景のみを含む画像を用いた。そのため本研究では真値データとして、注目しているブロックを「背景」、その他のすべてのブロックを「対象物体」とした(図10)。そして、注目しているブロックと、すべてのブロックを比較し、それぞれの場合について背景か対象かを推定して真値データと照らしあわせ、それらの結果をもとにROC曲線を描いた。

各ブロックにおいて、解像度を $M = 4, 8, 12, 16$ と

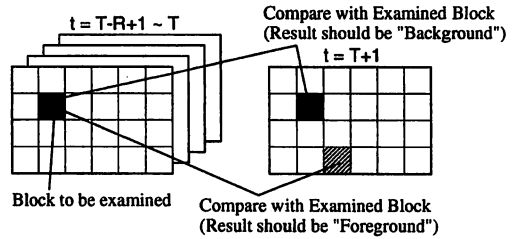


図10 対象領域の真値データ。

して背景差分を行い、ROC曲線を描いた。ただし、横軸は実際背景であるものを誤って対象物体と検出した割合、縦軸は実際対象物体であるものを正しく検出した割合を表す。各パターンのブロックについて、得られたROC曲線を図11に示す。

ROC曲線は、左上に位置するほど対象検出の精度が高いことを意味する。パターンAのブロックについては、解像度によらずROC曲線は高い検出精度を示したが、これは静止シーンの変動の予測が容易であるためと考えられる。パターンBでは $M = 4$ では精度はあまり良くないが、 $M = 8, 12$ の場合に高い精度を示し、 $M = 16$ でやや低下した。パターンCでは解像度を粗くするにつれて精度は向上するが、パターンBと同様 $M = 16$ でやや低下した。

この結果より、解像度が細かい場合 ($M = 4$) は予測誤差 E が大きいため検出精度は劣り、解像度を粗くするに伴い E が小さくなることで、おおむね検出精度は向上することがわかる。さらに極端に平滑化した場合 ($M = 16$) には検出精度はやや低下するが、これは極端な平滑化により状態ベクトルの変動値 V が小さく、背景と対象との識別能力が落ちるためと考えられる。以上の実験結果から次の3点が結論づけられる。

- 解像度を粗くしていくに従って対象検出の精度はおおむね向上していくが、極端に平滑化すると精度はそれ以上は向上せず、逆に低下することが確認できる。
- その精度を決定する要素として E と V が大きく関係しており、おおむね E が小さくなるに従い精度は向上するが、 E が小さくても V が極端に小さい場合は精度は悪化するといえる。
- したがって、対象検出に最適な状態ベクトルを決定する際、その解像度を E と V を評価値として決定できると考えられる。

5. 結論

本論文では、動的環境において対象検出を実現するために、線形システムを用いて動的背景の変動を記述する方法を提案した。さらに、画像の解像度などに着目し、対象検出を行う上で適切な状態ベクトルの決定

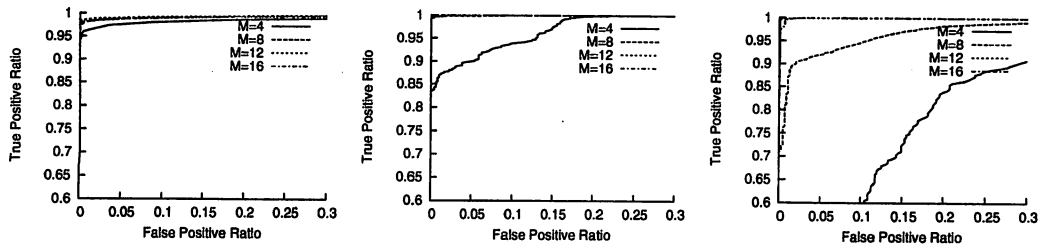


図 11 解像度に伴う検出精度の比較. 左から順にパターン A, B, C のブロック.

法を提案した.

本論文で提案した状態ベクトルの生成法を線形システムに適用することで動的背景の変動を精度よくモデル化でき、高精度の対象検出を実現できることを示した. さらに、状態ベクトルを選択するにあたって、状態ベクトルの予測誤差や変動値を評価値とすることで最適な解像度を決定することの妥当性を示した.

なお、以下の項目については本論文では十分に扱うことができなかったため、今後の課題とする.

最適な解像度の具体的な決定法: 対象検出に最適な状態ベクトルを選択する際、解像度を決定するために状態ベクトルの予測誤差や変動値を評価値とすればよいことが示された. しかし本論文ではその因果性が示せたに過ぎないので、予測誤差や変動値を求めたあと、解像度を決定するには閾値 Th_E をどの程度にするか、などを今後考察する必要がある.

周辺の画像ブロックとの関連性: 本論文では各画像ブロック単位で背景変動をモデル化したがる、各ブロックにおける背景変動は周辺のブロックとも何らかの共起性があると考えられる. 今後はそのような空間的な相関関係にも注目し、より正確な背景変動のモデル化を目指す.

不規則な背景変動の記述: パターン C のブロックのように静止部分と木の葉の部分が混在している場合、ブロック全体での変動は不規則であるために平滑化しても必ずしも線形な変動として表現できるとは考えにくいがる、本論文ではそれについて十分に議論することができなかった. このようなブロックにおける変動をどのように表現するかを今後考察する必要がある、たとえばブロックの分割や統合を行うことにより処理単位を変更する、などのアプローチが考えられる.

謝 辞

本研究の一部は、文部科学省プロジェクト「知的資産の電子的な保存・活用を支援するソフトウェア技術基盤の構築」における研究開発課題「大型有形・無形文化財の高精度デジタル化ソフトウェアの開発」の補

助を受けた.

参 考 文 献

- 1) 鷺見和彦, 関真規人, 波部齊. 物体検出-背景と検出対象のモデリング. 情報処理学会研究報告 (CVIM), pp. 79-98, 2005.
- 2) 長屋茂喜, 宮武孝文, 藤田武洋, 伊藤渡, 上田博唯. 時間相関型背景判定法による移動物体検出. 信学論 D-II, Vol. J79-D-II, No. 4, pp. 568-576, 1996.
- 3) B. Lo and S.A. Velastin. Automatic congestion detection system for underground platforms. In *International Symposium on Intelligent Multimedia, Video and Speech Processing*, pp. 158-161, 2001.
- 4) C.R. Wren, A. Azarbayejani, T.J. Darrell, and A.P. Pentland. Pfunder: Real-time tracking of the human body. *IEEE Trans. PAMI*, Vol. 19, No. 7, pp. 780-785, July 1997.
- 5) 波部齊, 大矢崇, 松山隆司. 動的環境における頑健な背景差分の実現法. 画像の認識・理解シンポジウム MIRU'98, Vol. I, pp. 467-472, Jul 1998.
- 6) 関真規人, 和田俊和, 藤原秀人, 鷺見和彦. 背景変化の共起性に基づく背景差分. 情報処理学会論文誌, Vol. 44, No. SIG 5, pp. 54-63, Apr 2003.
- 7) K. Toyama, J. Krumm, B. Brumitt, and B. Meyers. Wallflower: Principles and practice of background maintenance. In *ICCV99*, pp. 255-261, 1999.
- 8) 川嶋宏彰, 西山正敏, 松山隆司. 表情譜: タイミング構造に基づく表情の記述・生成・認識. 第 4 回情報科学技術フォーラム (FIT2005), pp. 153-156, 2005.
- 9) G. Doretto, A. Chiuso, Y. Wu, and S. Soatto. Dynamic textures. *International Journal of Computer Vision*, Vol. 51, No. 2, pp. 91-109, 2003.