

## 時空間テクスチャを用いた背景モデルによる 動体検知法

弓場竜<sup>†</sup> 三好雅則<sup>†</sup> 藤吉弘亘<sup>††</sup>

背景差分は動体検知の一般的な手法として広く適用されてきたが、背景に大きな変化が発生したときに、背景の変化と動体の進入を判別することが依然として課題である。そこで本稿では、動画像中の小領域内のアピランスの情報に加えてモーションの情報を反映した時空間テクスチャの Space-Time Patch 特徴量を用いることによって、背景の変化への追従性を高めた背景モデルによる動体検知の手法を提案する。また本稿では評価実験において、屋外で日照が変化し樹木が揺れるシーン、屋外に面したエスカレータから乗客が乗り出すシーン、屋内の天井照明が頻繁に明滅するシーンの3つのシーンにおいて、提案手法が従来手法より良好に動体を検知した結果を述べる。

### Moving Object Detection with Background Model based on Spatio-Temporal Texture

Ryo Yumiba<sup>†</sup> Masanori Miyoshi<sup>†</sup>  
Hironobu Fujiyoshi<sup>††</sup>

Background subtraction is a common method detecting moving objects, but it is yet a difficult problem to distinguish moving objects from background when the background changes seriously. Hence in this paper, we propose a method detecting moving objects with background model that covers dynamic changes in background utilizing spatio-temporal texture, "Space-Time Patch" which describes motion in addition to appearance like other conventional spatial texture. Our experimental results show the proposed method outperforms a conventional method in three scenes: in an outdoor scene in which tree branches flutter to wind, in an escalator scene besides a window facing outdoor in which people bodies stick outside, and in an indoor scene in which lightings on ceil change on and off frequently.

### 1. はじめに

背景差分は動体検知の手法として一般的であり、多くの監視システムに適用されてきた。背景差分による動体検知は、入力画像を事前に準備した背景モデルと照合して、入力画像のうち背景モデルに適合しない部分を動体として検知する手法である。この背景差分による動体検知には、動体の形状や運動等に事前知識を必要としない利点がある。一方で背景差分による動体検知には、背景に変化が生じた場合に背景の変化と動体の進入の判別が困難という課題がある。背景の変化には多様なバリエーションがあるが、大別すると日照の変化や天井の照明の切り替えのような画面上の大域的な変化と、風に揺らぐ樹木や水面の揺らぎなどによる画面上の局所的な変化がある。

これまでに背景の変化を吸収して動体を検知するための種々の背景モデルが提案されてきた。背景モデルの生成の1つ目のアプローチは、画像の最小単位の画素の変化を、過去の観測情報に基づいた統計分布でモデル化する方法である[1][2]。この画素を単位とした背景モデルには、草木や水面の揺らぎのような画像上で局所的に生じる頻繁な明度の変化を統計分布によって吸収できる利点がある。一方で、画素単位の背景モデルには、日照の強さの変化のような画像上の大域的な変化が短時間に発生すると、背景の統計分布が短い時間で変化してしまうために追従が困難という課題がある。

背景モデルの生成の2つ目のアプローチは、画像上の小領域内のテクスチャを照明の変化に頑健な特徴量でモデル化する方法である[3][4]。この小領域の単位の背景モデルには、日照の強さの変化の様な画像上の大域的な変化を、特徴量の頑健さによって吸収できる利点がある。一方で小領域の単位の背景モデルには、草木や水面の揺らぎのような画像上の局所的な変化が短時間に強く発生すると、小領域内のテクスチャが複雑に変化するために追従が困難という課題がある。

背景モデルの生成には、画素の単位と小領域の単位の両方の背景モデルを組み合わせた手法も提案されている[5]。この手法は、画素の単位と小領域の単位の検知結果の論理積を取って誤検知を抑止するものであるが、画面上の1つの場所を2通りの方法で信号処理することが冗長となる。

そこで本稿では、背景の変化への追従を高めた背景モデルとして、動画像中の小領域内の「アピランス（見え）」と同時に「モーション（動き）」の情報を合わせ持った時空間テクスチャを適用した背景モデルを提案する。この時空間テクスチャを適用した背景モデルは、小領域内のアピランスの情報をもって画像上の大域的な背景の変化に追従すると同時に、従来のアピランスの情報による小領域を単位とした背景

<sup>†</sup> 株式会社 日立製作所 日立研究所  
Hitachi, Ltd., Hitachi Research Laboratory  
<sup>††</sup> 中部大学大学院 工学研究科 情報工学専攻  
Department of Computer Science, Chubu University

モデルでは吸収が困難であった、画像上の局所的な背景の変化をモーションの情報で吸収するアプローチをとる。

本稿は以下、2章では本稿の提案手法で用いる時空間テクスチャの Space-Time Patch 特徴量について述べる。次に3章では、Space-Time Patch 特徴量から背景モデルを生成し、背景差分により動体を検知する方法について述べる。次に4章では、本稿の提案手法の評価実験について述べる。最後に5章では、本稿のまとめを述べる。

## 2. Space-Time Patch 特徴量

2章では、本稿の提案手法における時空間テクスチャの Space-Time Patch (以下、ST-Patch と略記) 特徴量について述べる。ST-Patch 特徴量は、Shechtman らが提案している動画像中の小領域内におけるアピランスとモーションを捉えた特徴量である [6]。この ST Patch 特徴量は、動画像中の動作領域の抽出 [6]のほか、イベントの検出 [7]や異常挙動の検出 [8]や動作識別[9]や移動方向別の物体抽出[10]等への適用例が報告されてきた。

ST-Patch 特徴量は、動画像の小領域における各画素の濃度の勾配値の統計量から求められる。動画像中のある画素の  $x$  軸,  $y$  軸の空間方向の濃度の勾配値  $(P_x, P_y)$ と,  $t$  軸の時間方向の濃度の勾配値  $P_t$  を合わせて時空間の濃度の勾配値  $(P_x, P_y, P_t)$ とすると, 小領域内の濃度の勾配値の2次元のモーメントは, 式(1)に示す行列  $M$  で表せる。

$$M = \begin{bmatrix} \sum P_x^2 & \sum P_x P_y & \sum P_x P_t \\ \sum P_y P_x & \sum P_y^2 & \sum P_y P_t \\ \sum P_t P_x & \sum P_t P_y & \sum P_t^2 \end{bmatrix} \quad (1)$$

式(1)の行列  $M$  は Gram 行列と呼ばれる。行列  $M$  の左上の時間の要素を含まない  $2 \times 2$  の要素は、動画像中の小領域内の画素のアピランスの傾向を反映する。行列  $M$  の間方向および時間方向と空間方向の両方の成分をあわせもった3行目と3列目の要素は動画像中の小領域の画素のモーションの傾向を反映する。ST-Patch 特徴量は、式(1)より求められる行列  $M$  の行列要素を構成要素とする。ただし、行列  $M$  は対称行列であるため、ST-Patch 特徴量は式(2)に示す行列  $M$  の重複しない行列要素の6次元のベクトルとなる。


$$v = [\sum P_x^2, \sum P_x P_y, \sum P_y^2, \sum P_x P_t, \sum P_y P_t, \sum P_t^2] \quad (2)$$

表 1 に ST-Patch 特徴量の例として、樹木の同じ部分の(1)静止した時、及び(2)に風

に揺れるとき、路面の同じ部分の(3)日照が安定したとき及び(4)日照が変化中の数値例を示す。小領域のサイズは横 15 画素×縦 15 画素×5 フレームとした。表 1 の ST-Patch 特徴量  $\{v_j\}_{j=1 \sim 6}$  には、下記の傾向がみとめられる。

- ・時間変化の有無によらず、樹木および路面でアピランスの要素は分離 ((1)と(2)および(3)と(4)の  $v_{1,2,3}$  の比較)
- ・時間変化があるとモーションの要素が増大 ((1)と(2)の  $v_{4,5,6}$ , (3)と(4)の  $v_{4,5,6}$  の比較)
- ・日照の変化と樹の揺れとで、モーションの要素が相異 ((2)と(4)の  $v_{4,5,6}$  の比較)

表 1 ST-Patch 特徴量の数値例



	$v_1$	$v_2$	$v_3$	$v_4$	$v_5$	$v_6$
	$\sum P_x^2$	$\sum P_x P_y$	$\sum P_y^2$	$\sum P_x P_t$	$\sum P_y P_t$	$\sum P_t^2$
(1)樹木が静止	68.3	1.1	38.4	-0.2	0.1	0.6
(2)樹木が風に動揺	66.5	-11.7	52.7	1.8	-1.3	1.7
(3)路面の日照が安定	0.7	-0.1	0.6	0.0	0.0	0.0
(4)路面の日照が変化中	1.4	-0.1	0.7	-0.4	0.0	1.2

アピランスの要素                      モーションの要素

## 3. Space-Time Patch 特徴量を用いた背景差分法

3章では ST-Patch 特徴量を使って、動的に変化する背景の背景モデルを生成して動体を検知する方法を述べる。まず 3.1 節では、背景の変化の状態が複数ある場合を想定して、背景モデルを ST-Patch 特徴量の混合正規分布により生成する方法を述べる。次に、3.2 節では入力画像の ST-Patch 特徴量を背景モデルと比較して、動体を検知する方法を述べる。最後に 3.3 節では、背景の変化の時間推移に追従するために、背景モデルを漸進的に更新する方法を述べる。

### 3.1 Space-Time Patch 特徴量と混合正規分布による動的な背景モデルの生成

背景の変化には、エスカレータのステップのように同一方向かつ一定速度の移動を継続する定常的なものもあれば、吹く風の強さに応じてランダムに揺れる樹木や雲の移動につれて変化する日照の強さのように非定常なものがある。また、背景の変化には、複数の変化が組み合わせて発生する状況も考えられる。そこで本稿の提案手法では、バリエーションを持った背景の変化を許容するために、式(2)で求めた ST-Patch 特徴量の混合正規分布により背景モデルを生成する。ここで、混合正規分布の数式を、式(3)および式(4)に示す。

$$p(x|C) = \sum_{i=1}^M w_i p(x|C_i) \quad (3)$$

$$p(x|C_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left\{-\frac{1}{2}(x-\mu_i)^T \sigma_i^{-1}(x-\mu_i)\right\} \quad (4)$$

式(3) および式(4)において、 $p(x|C)$ は確率変数  $x$  が混合正規分布  $C$  に属する確率密度、 $p(x|C_i)$ は確率変数  $x$  が  $i$  番目の正規分布  $C_i$  に属する確率密度、 $w_i$ ,  $\mu_i$ ,  $\sigma_i$  はそれぞれ、 $i$  番目の正規分布の重み係数、平均、標準偏差である。本稿の提案手法では、ST-Patch 特徴量の各要素  $\{v_j\}_{j=1\sim 6}$  それぞれが独立して、個別の混合正規分布に従うと仮定する。混合正規分布のパラメータ  $w_i$ ,  $\mu_i$ ,  $\sigma_i$  は、EM アルゴリズムを用いて事前に収集した背景のサンプル映像から計算しておく。

背景の変化は大半の場合において画像上の場所ごとに異なるので、以上述べた ST-Patch 特徴量の混合正規分布のパラメータは、画面上を所定サイズで均等に分割したブロック毎に計算する。

### 3.2 Space-Time Patch 特徴量による背景モデルを用いた動体検知

本稿の提案手法では入力映像が与えられた時、動画の各フレームから検知エリア内の ST-Patch 特徴量を抽出し、3.1 節で求めた背景モデルの混合正規分布から乖離したブロックを動体として検知する。以下に、本節の処理のステップを述べ、概要を図 1 に示す。

Step1. 入力映像の各フレームにおいて、検知エリア内の各ブロックの ST-Patch 特徴量を計算する。

Step2. Step1 の ST-Patch 特徴量の各要素  $\{v_j\}_{j=1\sim 6}$  を確率変数  $x$  とおいて、式(3) の  $x$  が背景モデルに属する確率密度と所定のしきい値  $\theta$  の大小を式(5)で比較する。

$$p(x|C) > \theta \quad (5)$$

Step3. ST-Patch 特徴量  $\{v_j\}_{j=1\sim 6}$  の全要素で式(5)が満たされれば、ST-Patch 特徴量は背景モデルに適合するとして、そのブロックを動体無と判定する。反対に、要素  $\{v_j\}_{j=1\sim 6}$  の中で一つでも式(5)が満たされない場合、ST-Patch 特徴量は背景モデルに適合しないとして、そのブロックを動体有と判定する。

Step4. Step3. で動体有と判定したブロックの数が、動体の大きさを想定したしきい値  $\pi$  個以上あれば、動体有の候補フラグを ON にする。

Step5. 動体有の候補フラグが ON のフレームが、フレーム毎の動体有の判定の確からしさより定めたしきい値  $\rho$  以上連続すれば、検知エリア内に動体有の発報を出力する。

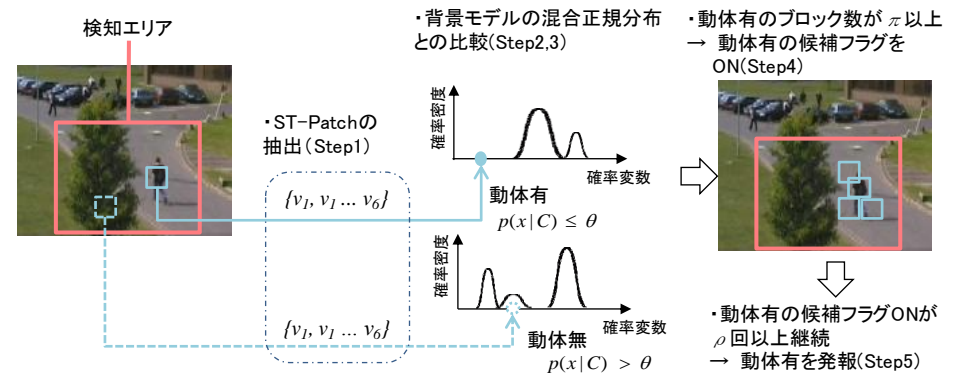


図 1 Space-Time Patch 特徴量による背景モデルを用いた動体検知の概要図

### 3.3 Space-Time Patch 特徴量による背景モデルの更新

背景が変化するシーンの一例として、屋外において日照条件が変化するシーンでは、背景の変化の様子が時間経過とともに推移する。また、風によってランダムに樹木が揺れるシーンでは、樹木の揺れ方のサンプルを事前にくまなく収集しておくことは困難である。そこで、本報告の提案手法では、時間経過による背景の変化の推移に追従していくために、動体検知を行う間に、背景モデルの混合正規分布を少しずつ更新する方法をとる。

背景モデルの更新において背景の変化が非定期的な場合、時間経過とともに変化が消滅すること、新たな変化が発生することが予測できる。また背景モデルの更新において、特性が近い変化は冗長であり 1 つに統合することが望ましい。そこで本報告の提案手法の背景モデルの更新には、混合正規分布中の正規分布に消滅/追加/統合の条件分岐を持たせた以下の手法を適用した[2]。図 2 に以下のステップの概要図を示す。

Step1. 入力映像の ST-Patch 特徴量  $\{v_j\}_{j=1\sim 6}$  が背景に属するか否かを、3.2 節の Step2 の方法で判定する。

Step2a. Step1 で背景に属すと判定した場合、確率変数  $x$  にもっとも近い  $i$  番目の正規分布の重み係数  $w_i$ , 平均  $\mu_i$ , 標準偏差  $\sigma_i$  を、所定の更新率  $\alpha$  と式(6)および式(7)で更新する。

$$\begin{aligned} w_i' &= (1-\alpha)w_i + \alpha \\ \mu_i' &= (1-\eta)\mu_i + \eta x \\ \sigma_i'^2 &= (1-\eta)\sigma_i^2 + \eta(x-\mu_i)^T(x-\mu_i) \end{aligned} \quad (6)$$

$$\eta = \alpha p(x | C_i) \quad (7)$$

Step2b Step1 で背景に属しないと判定した場合、正規分布の数  $M$  が許容値  $M_{max}$  未満であれば、平均を  $x$ 、重み係数と標準偏差を所定値  $w_o$ 、 $\sigma_o$  とした正規分布を混合正規分布に追加する。

- Step3 以下の基準で、微小なあるいは冗長な正規分布を削除する。
- 重み係数が所定値  $w_{min}$  未満であれば、ノイズとみなして削除する。
  - 2つの正規分布の平均が近い場合、片方を削除して1つに統合する。

Step4 正規分布の重み係数  $\{w_i\}$  の総和が1になるように正規化する。

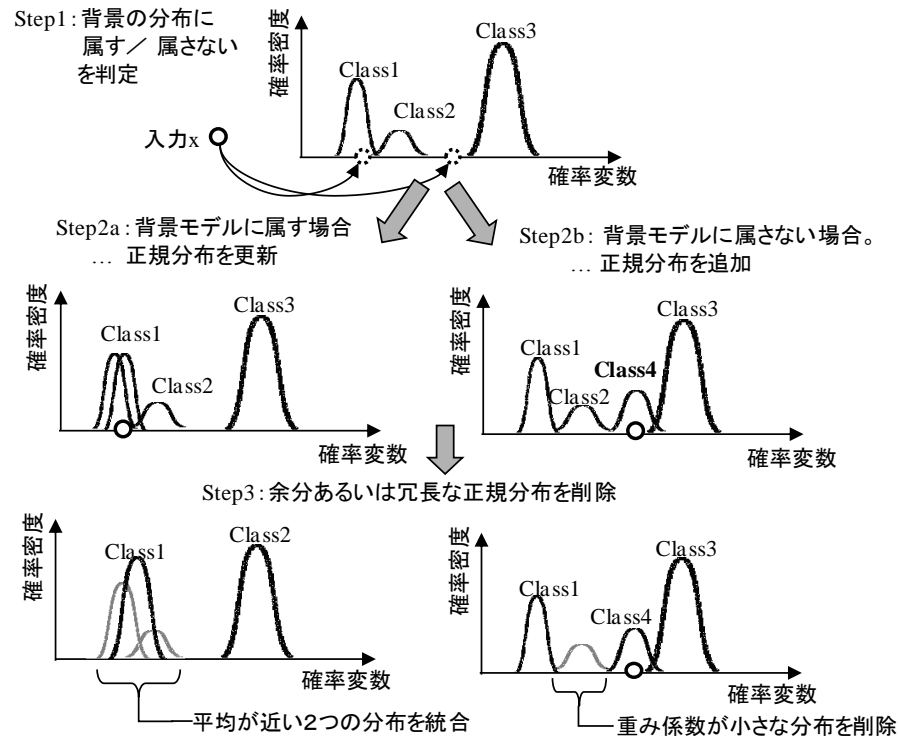


図 2 Space-Time Patch 特徴量による背景モデルの更新の概要

## 4. 評価実験

### 4.1 実験概要

4章では、3つのシーンにおける本稿の提案手法の評価実験を述べる。評価実験では、アピアランスの特徴量を用いた従来手法として、正規化距離[4]を濃淡画像に適用した動体検知を同時に評価した。なお、正規化距離による動体検知のアルゴリズムは、3.2節で述べた検知エリアの範囲、検知エリア内のブロックの分割、およびブロック単位で動体の有無を判定した後段の処理(3.2節の Step4 および Step5)は、ST-Patch 特徴量による方法と条件を共通にした。また、正規化距離による動体検知では、背景生成用の映像の平均画像を背景画像として、各画素を式(7)と同じ更新率  $\alpha$  による IIR フィルタで更新した。なお、更新率は4章の評価実験を通じて 0.01 とした。

その他のパラメータとして、映像のサイズは横 320 画素×240 画素、フレームレートは 30fps に対して、ST-Patch 特徴量のブロックのサイズは横 15 画素×縦 15 画素×5 フレーム、正規化距離のブロックのサイズは横 15 画素×縦 15 画素とした。また、ST-Patch 特徴量の混合正規分布の正規分布の数は、最大で3つとした。

定量評価の指標としては、まず評価映像の中で動体が十分な大きさで映っているフレームを動体有、それ以外のフレームを動体無とし、フレーム単位で動体の有無と発報の有無を比較して、TP (動体有で発報有)、TN (動体無で発報無)、FP (動体無で発報有)、FN (動体有で発報無) に分類した。次に、全フレームの比較結果の集計から、式(8)で計算した Recall (再現率) と Precision (適合率) を定量評価の指標とした。

$$\text{Recall} = \frac{\sum \text{TP}}{\sum \text{FN} + \sum \text{TP}} \quad (8)$$

$$\text{Precision} = \frac{\sum \text{TP}}{\sum \text{FP} + \sum \text{TP}}$$

加えて、Recall と Precision の総合評価として、F-Measure (Recall と Precision の調和平均) を定量評価の指標とした。

### 4.2 屋外のシーンにおける評価実験

1つ目の評価実験では、PETS2001 で公開された画面の中央部に樹木がある屋外のシーンの評価映像を対象とした。この PETS2001 の評価映像では間断的に風が吹き、画面の中央部の樹木は吹く風につれて非定常的に揺れて局所的な変化になっている。また、この PETS2001 の評価映像では上空の雲の動きにつれて、日照の強度が短時間の間に大きく変化し、画像上の大域的な明度が大きく変化する。4.2節の評価実験では、この PETS2001 の屋外のシーンの中で、特に背景の変化が激しい樹木の付近を検知エリアに設定した。

PETS2001 の屋外映像には、学習用と評価用の2つの映像が用意されているので、

最初に学習用の映像から検知エリア内に人物がない 6668 フレームを抜粋して、背景モデルを生成した。図 3 に、4.2 節の評価実験で背景生成に用いた映像のサンプルのフレームを示す。



図 3 屋外のシーンの背景生成に用いた映像のサンプル

次に、PETS2001 の屋外のシーンの評価用の映像で、動体検知の性能を評価した。検知の対象とした動体は、樹木の近くを通る通行者である。図 4 に動体検知の結果の代表例を示す。図 4 の 2 列目と 3 列目において、画像の中央の樹木にまたがる四角形の描画は検知エリア、検知エリア内の小さいブロックの描画は動体有と判定したブロックを示す。また、図 4 の 2 列目と 3 列目の画像の左上の角には、そのフレームの動体および発報の有無を記した。

図 4 の(1)は、風が強くなり樹木の揺れが徐々に大きくなりだしたフレームである。図 4 の(1)において、正規化距離による方法は樹木の一部に動体有のブロックが出て誤報しているが、ST-Patch 特徴量による方法は動体有のブロックがほとんど誤報していない。図 4 の (2)は図 4 の(1)の後に風が強くなり、背景の生成時よりも樹木が強く揺れ出したフレームである。図 4 の(2)では、正規化距離と ST-Patch 特徴量のいずれの方法も、動体有のブロックが多数出て誤報している。図 4 の(3)は、図 4 の(2)の後も強い樹木の揺れが継続した 289 フレーム後である。図 4 の(3)において正規化距離による方法は多数の動体有のブロックが多数あり依然として誤報している一方で、ST-Patch 特徴量による方法は樹木の揺れに追従して、動体有のブロックがなくなり誤報していない。図 4 の(4)は、図 4 の(3)から後に日照が弱くなったフレームである。図 4 の(4)では、正規化距離による方法は画面全体に動体有のブロックが出て誤報している一方で、ST-Patch 特徴量による方法は動体有のブロックがほとんどなく、樹木および画像全体の日照の変化に追従して誤報を抑制している。

図 4 の代表例に示す以外の傾向としては、ST-Patch 特徴量による方法は人物が樹木に近いと失報する傾向が見られた。また正規化距離による方法は、人物が樹木に近いとき、および人物の身体の一部の濃淡値が路面に近いとき失報する傾向が見られた。

4.2 節の評価実験では、評価映像の 6255 フレームのうち、通行者が十分な大きさで

映っている 1359 フレームを動体有、それ以外のフレームを動体無とした。表 2 に 4.2 節の定量評価の結果を示す。表 2 より、4.2 節の屋外のシーンにおいて、ST-Patch 特徴量による方法が正規化距離による方法よりも、Recall で 16%と Precision で 38%と両方で上回り、総合的な F-Measure で 32%と大幅に上回ったことがわかる。



図 4 屋外のシーンの検知結果の例

表 2 屋外のシーンの評価結果

	Recall (%)	Precision (%)	F-Measure (%)
正規化距離	76	41	53
ST Patch	92	79	85

### 4.3 エスカレータのシーンにおける評価実験

2つ目の評価実験では、実験設備のエスカレータにおいてハンドレールから乗り出す乗客の身体を動体検知の対象とした。このハンドレールから乗り出した乗客の身体は、乗客の動作とともに、エスカレータのステップの運行によって画面上を移動する。4.3節の評価実験で用いる実験設備のエスカレータは、屋外に面した大きな窓が傍にあるために、差し込む日照の強さが時間とともに推移する。この日照の強さの変化によってハンドレールおよびハンドレールの外側の金属部の明度および明度むらは変化し、これが4.3節のシーンの大局的な変化となる。また4.3節の評価実験では、ステップに乗客が乗車すると窓を通した外光により乗客の影が発生し、ハンドレール外側の金属部に差し込む。この乗客の影は、4.3節のシーンの局所的な変化となる。

4.3節のシーンでは、ハンドレールから乗り出さない乗客が1人通過する映像から背景モデルを生成した。図5に、背景モデルを生成した映像のサンプルを示す。

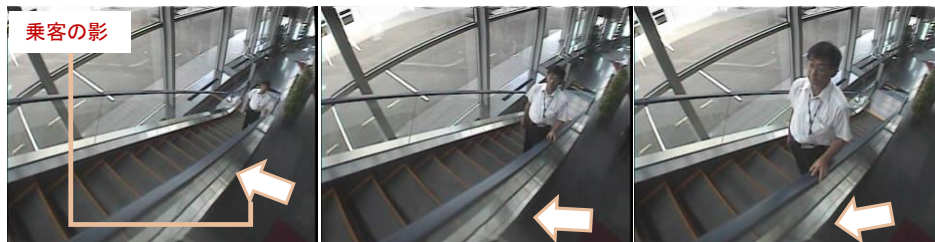
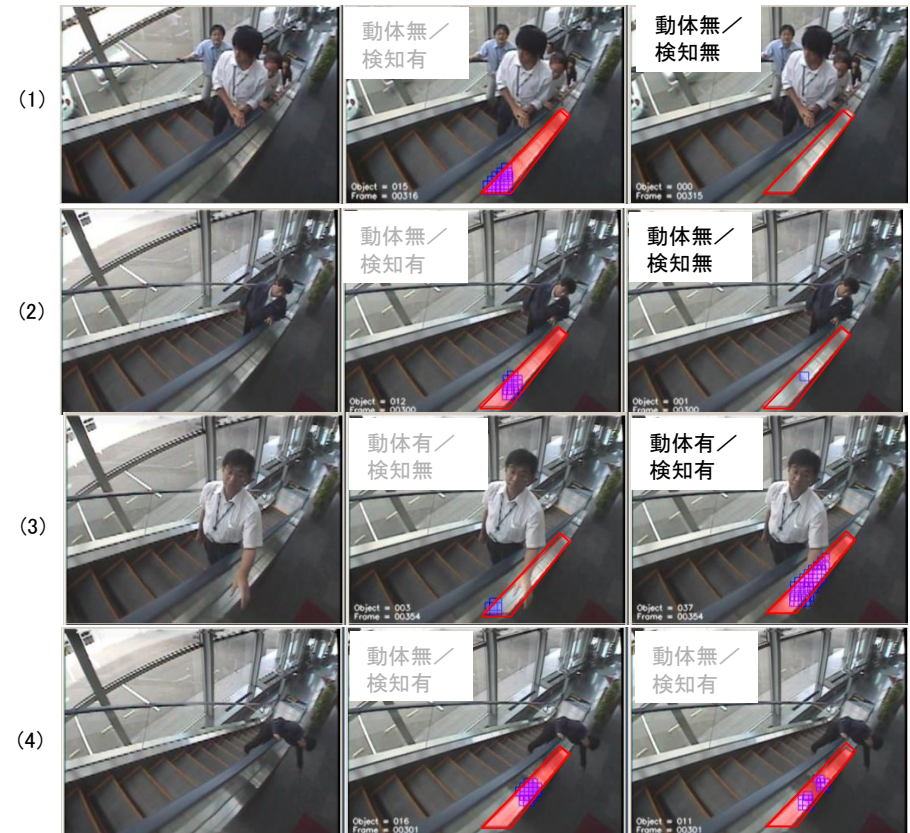


図 5 エスカレータのシーンの背景映像のサンプル

次に、撮影実験で被験者がエスカレータから乗り出す挙動を模擬した映像を使って、動体検知を評価した。被験者のエスカレータから乗り出す挙動には、乗り出す部分を頭部、腕部、上部のようにバリエーションを持たせた。また、影の向きに違いができるように、乗客の人数を複数人の場合と1人の場合のケースで行った。

図6に、4.3節の評価実験の結果の代表例を示す。図6において、帯状の四角形はハンドレールの外側の金属部に設けた検知エリア、検知エリア内の小さいブロックは動体有と判定したブロックを示す。



(a)入力画像 (b)正規化距離の結果 (c)ST-Patchの結果

図 6 エスカレータのシーンの動体検知の結果の例

図6の(1)は複数の乗客が乗り出すことなく通過しているフレームである。また、図6の(2)は乗客が検知エリア付近に到達する以前のフレームである。図6の(1)と(2)において、正規化距離による方法は背景の生成時の乗客の影と動体の検知時の影の向き違いを吸収できず、影の付近に動体有のブロックが発生して誤報している一方で、ST-Patch 特徴量による方法では動体有のブロックはほとんど発生せず誤報していない。図6の(3)は、乗客の腕がはみ出したフレームである。図6の(3)において、正

規化距離による方法は肌と検知エリア内の金属部のコントラストが低いために失報した一方で、ST-Patch 特徴量による方法は乗客の腕付近に動体有のブロックが発生して発報している。図 6 の(4)は、乗客の上半身が大きくはみ出したフレームであり、はみ出した乗客が窓からの光を遮って濃い大きな影ができています。図 6 の(4)では、背景モデルの生成に用いた図 5 の乗り出さない乗客の影と、図 6 の(4)で大きく乗り出した乗客の影とで見え方が大幅に変わっているために、正規化距離による方法も ST-Patch 特徴量による方法も背景モデルが吸収できずに動体有のブロックが発生し、誤報している。図 6 の代表例以外の傾向として、ST-Patch 特徴量による方法は、人物の身体が検知エリアへ進入した初期のフレームで失報する傾向がみられた。ただし、ST-Patch 特徴量による方法では、検知エリアへの進入の初期のフレームで失報になった乗客はすべて、以降のフレームで 1 回以上発報されていた。

表 3 に、評価映像の 3350 フレームを対象とした 4.3 節のシーンの定量評価の結果を示す。なお、評価映像において乗客の身体が十分な大きさで検知エリアに進入した 215 フレームを動体有のフレームとした。表 3 の 1 行目と 2 行目を比較すると、4.3 節のエスカレータのシーンにおいて、ST-Patch 特徴量による方法が正規化距離による方法よりも、Recall で 12%と Precision で 36%と両方で上回り、総合的な F-Measure で 32%と大幅に上回ったことがわかる。

表 3 エスカレータのシーンの評価結果

	Recall (%)	Precision (%)	F-Measure (%)
正規化距離	78	34	47
ST Patch	90	70	79

#### 4.4 天井照明を明滅させるシーンの評価実験

3 つ目の評価実験では、実験室内の天井照明を頻繁に明滅させるシーンの公開映像 [11]において、画面を横切る人物を動体検知の対象とした。評価対象とした公開映像では実験室内の天井に複数の照明があり、天井照明の明滅があると画面中の明度は数フレームの短期間に急峻に変化する。4.4 節のシーンでは、画面の全体的な明るさは点灯している天井照明の数に応じて変化し、これが大域的な変化となっている。一方、4.4 節のシーンにおける画面中の個別の物体の明るさは、点灯している照明の場所と物体表面の法線方向の位置関係に応じて変化し、これが局所的な変化となっている。

4.4 節のシーンでは、公開映像の前半 665 フレームの人物が進入しない部分から背景モデルを生成した。図 7 に背景モデルを生成した映像のサンプルを示す。図 7 において画面の中央部には棒状の立体物があり、天井照明の変化に応じて特に強く明度が変化している。



図 7 天井照明を明滅させるシーンの背景生成に用いた映像のサンプル

次に、公開映像の後半 1981 フレームを評価映像に使用して動体検知を評価した。4.4 節のシーンでは、明度の変化が顕著な画面の中央部の棒状の立体物の周囲に検知エリアを設けた。この評価映像の中で、検知エリア内に人物が十分な大きさで映っている 99 フレームを動体有とした。

天井照明を明滅させるシーンの動体検知の結果の代表例を図 8 に示す。図 8 において、画面中央の四角形は検知エリア、検知エリア内の小さいブロックは動体有と判定したブロックを示す。

図 8 の(1)は画面中央の付近を照らす天井照明以外が消灯した直後である。図 8 の(2)は図 8 の(1)で消灯していた天井照明を全て点灯した直後である。図 8 の(1)と(2)において、正規化距離による方法では棒状の立体物の付近のブロックを中心に動体有と判定して誤報しているが、ST-Path 特徴量による方法では動体有と判定したブロックはわずかで誤報していない。図 8 の(3)は、点灯する天井照明が少なく暗い状況で、暗い服のコントラストが低い人物が横切った場面である。図 8 の(3)において正規化距離による方法では、動体有と判定したブロックが少なく失報しているが、ST-Patch 特徴量による方法は、人物付近の多くのブロックを動体有と判定して発報している。図 8 の代表例以外の傾向として、ST-Patch 特徴量による方法は、全 8 回の天井照明の明滅のうち 2 回で、明滅の直後の数フレームの間だけ誤報することがあった。一方、正規化距離による方法は、全 8 回の天井照明の明滅のうち 5 回で、明滅の直後およびしばらくの間に誤報していた。

表 4 に 4.4 節シーンの定量評価の結果を示す。表 4 の 1 行目と 2 行目を比較すると、4.4 節の天井照明を明滅させるシーンにおいて、ST-Patch 特徴量による方法が正規化距離による方法よりも、Recall で 27%と Precision で 58%と両方で上回り、総合的な F-Measure で 55%と大幅に上回ったことがわかる。

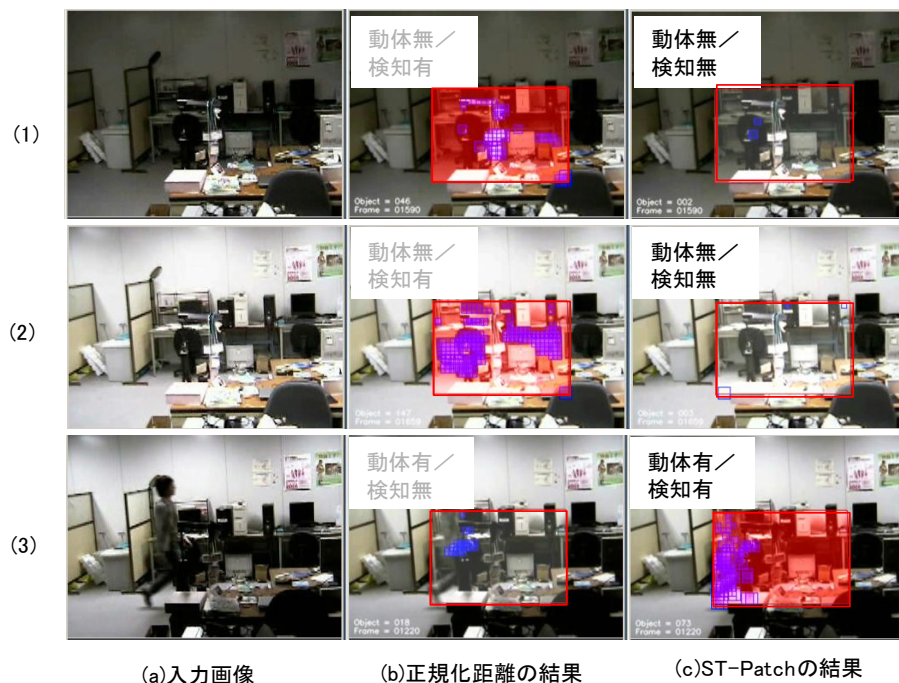


図 8 天井照明を明滅させるシーンの結果の例

表 4 天井照明を明滅させるシーンの評価結果

	Recall (%)	Precision (%)	F-Measure (%)
正規化距離	64	26	37
ST Patch	91	94	92

## 5. おわりに

本稿では、アピランスとモーションの情報を併せ持った時空間テクスチャの ST-Patch 特徴量を用いた背景モデルによる動体検知の手法を提案した。また評価実験において、モーションの情報を持たないアピランスの情報のみを用いた従来手法の一例に挙げた正規化距離との実験結果の比較より、提案手法が画面上の局所のおよび

大域的な背景の変化を吸収して良好に動体を検知する有効性を示した。

ただし、本稿の提案手法は検知エリアを単位とした動体検知であるために、動体に関する情報を抽出できない。本稿の提案手法を発展させて、背景が変化する環境において動体の位置や大きさや動線等の情報を頑健に抽出することを今後の課題に挙げる。

**謝辞** 本稿のアルゴリズムの研究開発およびエレベータのシーンでの実験評価に多大な貢献をいただいた数井誠人氏と村井泰裕氏に謹んで感謝の意を表す。また、本稿のエスカレータの評価映像の収集にご協力いただいた、株式会社日立ビルシステムの山口伸一郎氏と山崎浩氏に謹んで感謝の意を表す。

## 参考文献

- 1) A. Elgammal, D. Harwood, L. Davis: "Nonparametric Model for Background Subtraction", 6th European Conference on Computer Vision (ECCV), 2000
- 2) 島田敬士, 有田大作, 谷口倫一郎: "混合ガウス分布による動的背景モデルの分布数増減法", 画像の認識・理解シンポジウム(MIRU), pp.746-751, July 2006
- 3) 佐藤雄隆, 金子俊一, 丹羽義典, 山本和彦: "Radial Reach Filter(RRF) によるロバストな物体検出", 信学論 D-II, J86-D-II, pp. 616-624 (2003)
- 4) 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
- 5) 田中達也, 島田敬士, 谷口倫一郎, 山下隆義, 有田大作: "時空間特徴を考慮した動的背景モデル構築とそれに基づく物体検出", 画像の認識・理解シンポジウム(MIRU), pp.127-134, July 2009
- 6) Shechtman and M. Irani, "Space-Time Behavior Based Correlation," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2005
- 7) Y.Ke, R.Sukthankar and M.Hervert: "Event Detection in Crowded Videos", In Proc. ICCV2007, pp.8-15, 2007
- 8) 村井泰裕, 藤吉弘亘, 数井誠人: "時空間特徴に基づくエスカレータシーンにおける人の異常行動検知", パターン認識・メディア理解研究会(PRMU), pp.247-254, Sep 2008
- 9) Masato Kazui, Masanori Miyoshi, Shoji Muramatsu: "Incoherent Motion Detection using a Time-series Gram Matrix Feature". 19th International Conference on Pattern Recognition(ICPR), pp.1-5, Dec. 2008
- 10) 村井泰裕, 藤吉弘亘, 金出武雄: "Space - Time Patch を用いた物体の移動方向識別とセグメンテーション", コンピュータビジョンとイメージメディア(CVIM), Vol.1 No.2, pp.21-31, July 2008
- 11) 公開映像データ: <http://limu.is.kyushu-u.ac.jp/dataset/>