

時空間特徴の統計的学習を用いた異常行動検知

弓場 竜^{†a)} 村井 泰裕^{††b)} 藤吉 弘亘^{†††c)}

Abnormal Motion Detection Using Spatio-Temporal Features and Statistical Learning

Ryo YUMIBA^{†a)}, Yasuhiro MURAI^{††b)}, and Hironobu FUJIYOSHI^{†††c)}

あらまし 本論文では、動画像中の局所的な見えと動きの情報を併せ持った Space-Time Patch 特徴量と、統計的学習の Real Adaboost を用いて、動的な背景下の人の突発的な異常行動を検知する手法を提案する。提案手法は、時空間特徴の統計的学習による人領域の検出、及び統計的学習の信頼度を用いて人領域内の人の部分に重み付けした非定常度による異常行動検知の二つの要素から構成される。提案手法は、見かけの動きが小さな異常行動に対して、立体高次局所自己相関を適用した従来手法よりも有効である。ステップが動くエスカレータを背景、エスカレータの異常行動で代表的な転倒を対象とした評価実験において、提案手法は従来手法と比較して誤検知率 10% 時の検知率を 16% 向上させることができた。

キーワード Space-Time Patch, 人領域検出, 異常行動検知

1. ま え が き

近年、大規模な監視カメラの導入とともに、移動物や滞留物の検出等の動画像の認識技術によるセキュリティシステムの要求が高まりつつある。現在、これらの認識技術の発展として、不審行動や事故等の異常行動検知の研究が数多く進められている。異常行動の自動的な検知が実現すれば、監視者の負担軽減や人為的ミスの回避が可能である。これは、セキュリティシステムのコスト削減や異常の早期発見に寄与する。

異常行動検知の先行研究には、Scholkopf らの SVM を用いた確率分布の推定による例外値を検出する手法 [1] や、Zhong らの動画像の特徴の共分散行列を用いた手法 [2] 等がある。

先行研究の中で最も代表的な手法には、南里らが

提案する立体高次局所自己相関 (Cubic Higher-order Local Auto-Correlation) (以下 CHLAC) を用いた手法 [3] がある。CHLAC とは 2 値化したフレーム間差分の時系列に局所的な変化のパターンを当てはめて積算した特徴量である。この CHLAC は見えと動きの情報を併せ持つ。南里らの手法は、CHLAC と部分空間法の組合せで異常行動を検知する。まず学習のフェーズでは、正常行動のサンプルから CHLAC の部分空間を構築する。次に検知のフェーズでは、入力映像から算出した CHLAC の部分空間からの逸脱度を求め、逸脱度がしきい値以上の場合に異常行動として検知する。この方法は、個別の人の切出しが不要でロバストである。CHLAC は異常行動のほか、各種のイベントシーンからのイベントの検知にも用いられている [4]。

ほかにも、フローベクトルを使った方法も先行研究の中で代表的である。関らは、エレベータ内においてフローベクトルを算出し、その個数並びに方向と強度の分散から暴れの異常行動を検知する手法 [5] を提案している。都築らの手法 [6] は、SIFT 特徴量による特徴点の抽出と、特徴点のフローベクトルの抽出と、正常時のフローベクトルの方向のモデル化によって、モデルを逸脱する動きを異常行動として検知する。

異常行動検知をより多種類のシーンに適用するため

[†] (株)日立製作所日立研究所, 日立市
Hitachi, Ltd., Hitachi Research Laboratory, 7-1-1 Omika,
Hitachi-shi, 319-1292 Japan

^{††} (株)日立情報制御ソリューションズ, 日立市
Hitachi Information & Control Solutions, Ltd., 5-1-26
Omika, Hitachi-shi, 319-1221 Japan

^{†††} 中部大学大学院工学研究科情報工学専攻, 春日井市
Department of Computer Science, Chubu University, Chubu
University, 1200 Matsumoto, Kasugai-shi, 487-8501 Japan

a) E-mail: ryo.yumiba.xp@hitachi.com

b) E-mail: yasuhiko.murai.jw@hitachi-ics.co.jp

c) E-mail: hf@cs.chubu.ac.jp

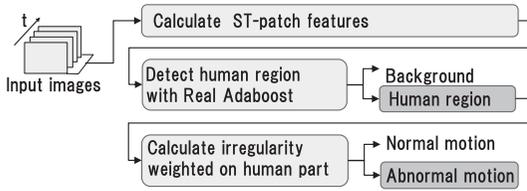


図1 提案手法の流れ

Fig. 1 Overview of the proposed method.

には、背景中の物体が動く動的な背景への対応が必要である。CHLAC やフローベクトルを用いた従来の手法は、この動的な背景への対応は困難である。背景の動きの特微量の中に人の動きの特微量が埋もれてしまい、人の異常行動の検知が困難である。この問題は、異常行動の見かけの動きが小さいほど顕著となる。

本論文は、異常行動の中で特に突発的なものを対象とし、動的な背景下において異常行動を検知する手法を提案する。突発的な異常行動とは、歩行中の転倒や階段の昇降中の踏み外しのような正常行動から逸脱して起きる異常行動である。提案手法はベースとして、映像中の物体の見えと動きの情報を併せ持った時空間特徴の適用により、背景と人の見えと動きの特徴抽出及び動きの非正常度の計算を行う。提案手法はこのベースに加えて、時空間特徴の統計的学習による人領域の検出と、非正常度の計算における統計的学習の信頼度による重み付けを行う。前者は方形単位で、背景と人を大まかに区分する。後者は、人領域内の非正常度の計算において、人領域内の人の部分に重み付けをすることで、背景の動きの影響を抑止する。提案手法は、この二つの要素で、動的な背景から突発的な異常行動を検知する。図1に提案手法の流れを示す。

本論文では、2.にて提案手法で用いる Space Time Patch 特微量（以下 ST-patch 特微量）について述べる。3.にてこの特微量を用いた異常行動検知の方法を述べる。4.にて、エスカレータを動的な背景、転倒を異常行動とした評価実験を行い、提案手法の有効性を示す。5.にて、本論文のまとめを述べる。

2. ST-patch 特微量

本章では、動画像の見えと動きを特徴記述する ST-patch 特微量について述べる。また、ST-patch 特微量から動きの非正常度を計算する方法を述べる。

2.1 ST-patch 特微量による動画像の特徴抽出

ST-patch 特微量は、Shechtman らが提案する動画

像の小領域内の見えと動きを捉えた特微量である [7]。この ST-patch 特微量は、特定動作の抽出 [7]、イベントの検出 [8]、動作の識別 [10]、移動方向別の物体検出 [11]、カット編集点の推定 [12]、シーン境界の検出 [13] 歩行状態の識別 [14] 等へ適用されてきた。

ST-patch 特微量は、動画像の小領域内の各画素の濃度のこう配値の統計量から求められる。x, y, z 軸をもつ動画像において、ある画素の空間方向のこう配値 (P_x, P_y) と、時間方向のこう配値 P_t を合わせて時空間のこう配値 (P_x, P_y, P_t) とする。このとき、画素数 n の小領域内のこう配値のスタックは式 (1) の行列 G で表せる。

$$G = \begin{bmatrix} P_{x1} & P_{y1} & P_{t1} \\ P_{x2} & P_{y2} & P_{t2} \\ \cdots & \cdots & \cdots \\ P_{xn} & P_{yn} & P_{tn} \end{bmatrix}_{n \times 3} \quad (1)$$

式 (1) から、動画像の小領域内のこう配値の二次元のモーメントは式 (2) の行列 M で計算できる。

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

行列 M の左上の 2×2 の時間方向のこう配値を含まない要素は、動画像中の小領域内の見えを反映する。行列 M の 3 行目と 3 列目の空間方向及び時間方向のこう配値をもつ要素は、動画像中の小領域内の動きを反映する。ST-patch 特微量は行列 M の重複しない行列要素を構成要素とする。行列 M は対称行列なので、ST-patch 特微量は式 (3) の 6 次元のベクトルとなる。

$$\nu = \left[\sum P_x^2, \sum P_x P_y, \sum P_y^2, \sum P_x P_t, \sum P_y P_t, \sum P_t^2 \right] \quad (3)$$

提案手法は式 (3) の ν のノルムを 1 に正規化する。これにより照明の変動による ν の強度の変動を抑止する。

2.2 ST-patch 特微量による非正常度の抽出

正常行動をとっていた人が突発的に異常行動をとると、時空間のこう配値は非正常な変化をする。ST-patch 特微量からは、Continuous Rank-Increase Measure

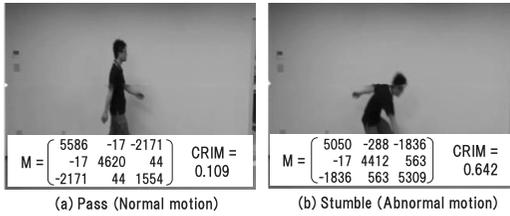


図 2 CRIM の数値例
Fig. 2 Example values of CRIM.

(以下 CRIM) [7] という時空間のこう配値の非定常度の指標が計算できる。提案手法では、この CRIM を異常行動の指標とする。

CRIM は以下の手順で計算する。今、式 (2) の行列 M の固有値を $\lambda_1 \geq \lambda_2 \geq \lambda_3$ とし、行列 M の上 2×2 を行列 M^\diamond とし、行列 M^\diamond の固有値を $\lambda_1^\diamond \geq \lambda_2^\diamond$ とすると、 $\lambda_1 \geq \lambda_1^\diamond \geq \lambda_2 \geq \lambda_2^\diamond \geq \lambda_3$ の関係にある。これらの固有値から、CRIM は式 (4) の Δr で計算する。

$$\Delta r = \lambda_2 \lambda_3 / \lambda_1^\diamond \lambda_2^\diamond \quad (0 \leq \Delta r \leq 1) \quad (4)$$

照明の変化が無視できるとき、 Δr は小領域内の動きの非定常度を示す。物体が静止若しくは一律な並行移動を継続する場合は、 Δr は 0 に近づく。反対に、物体の動きが変化する場合、 Δr は 1 に近づく。

図 2 の例を用いて、CRIM に正常行動の継続と異常行動の発生を区分できる特質があることを示す。この例では画像全体から CRIM を計算した。図 2 (a) は画面の左端から右端まで通過した正常行動である。図 2 (b) は左端から右端に通過する途中で転倒した異常行動である。 M 行列の 3 列の動きの要素を比較すると、後者は前者より 2 行と 3 行の値が大きい。CRIM は前者が 0.109、後者が 0.642 であり、後者の方が大幅に大きい。

3. ST-patch 特徴量と統計的学習を用いた異常行動検出

本章では、ST-patch 特徴量の統計的学習を用いて、背景の動きの影響を抑止して異常行動を検知する方法を述べる。まず ST-patch 特徴量の統計的学習を用いて人領域を検出する方法を述べる。次に人領域内の信頼度から、人領域内の人の部分に重みを与えた CRIM で異常行動を検知する方法を述べる。

3.1 統計的学習による人領域の学習と検出

近年、局所特徴量と統計的学習を用いた物体検出の方法が多く提案されている。Haar 状特徴量の統計的

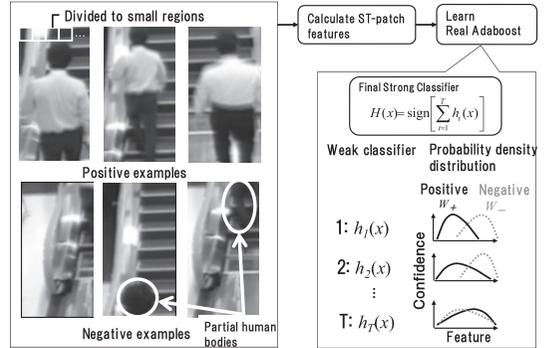


図 3 人領域の検出の統計的学習の概要
Fig. 3 Overview of statistical learning of human region detection.

学習による顔領域の検出 [16] や、Histograms of Oriented Gradients 特徴量の統計的学習による人領域の検出 [17] が代表例である。提案手法はこれらの手法と同様に人領域の検出に ST-patch 特徴量の統計的学習を用いる。

提案手法では、人領域の学習と検出に Real Adaboost [15] という手法を用いる。この手法は学習時に、特徴量の要素ごとに Positive と Negative の両クラスの確率密度関数を求め、両者の分離度が高いものを所定の個数だけ選択する。個々の特徴量に対応した識別関数は弱識別器と呼ばれる。最終的な識別結果は、弱識別器の識別結果の和となる。これは強識別器と呼ばれる。図 3 に人領域の検出の統計的学習の概要を示す。提案手法では、Positive クラスを正常行動時の人の領域、Negative クラスを背景及び人を部分的に捉えた領域とした。Negative クラスに人を部分的に捉えた領域を含めた目的は、通過する人の周辺で CRIM が増加することの抑止である。ここで、通過する人の周辺の局所に着目すると、人の進入若しくは退出は時空間の突発的な変化である。人の周辺を人領域として誤検出すると、その領域内の CRIM はこの変化から高い値をとってしまう。高い値の CRIM は、3.2 で述べる異常行動の検知に悪影響するので抑止をした。

人領域の検出は、学習結果を用いて次のように行う。まず、入力映像の各フレームを小領域の集合に分割して、小領域ごとの ST-patch 特徴量を抽出する。次に、検出ウィンドウをラスタ走査し、検出ウィンドウ内の小領域の ST-patch 特徴量から求めた Real Adaboost の強識別器の出力がしきい値を超えたものを人領域として検出する。最後に、検出した人領域は、Mean-Shift

のクラスタリングで近接したものを統合する。

提案手法において人領域の検出のパラメータは、以下のようにした。検出ウィンドウのサイズは 70×130 [pixel]、検出ウィンドウ内の小領域のサイズは 13×13 [pixel] $\times 5$ [frame] 及び間隔は 3 [pixel] とした。検出ウィンドウ内の小領域の個数は 741 個、特徴量の全次元は $4446 (= 741 \times 6)$ である。

3.2 統計的学習の信頼度を重み付けた非定常度の計算

提案手法では背景の動きの影響を抑止するため、人領域内の人の部分に重み付けした時空間のこう配値から CRIM を計算する。この重み付けには、人領域内の小領域ごとの Negative (背景) に対する Positive (人) の統計的学習の信頼度を用いる。この信頼度が大きな小領域ほど、人の部分を捉えている可能性が高いと考えて大きな重みを与える。

統計的学習の信頼度は、まず弱識別器ごとに求め、次に小領域ごとに求める。弱識別器の信頼度は Real Adaboost の応答値と分離度の二つの指標で計算する。

指標の一つ目の応答値の計算の方法を述べる。Real Adaboost の弱識別器はそれぞれ、Positive 及び Negative クラスの確率密度関数をヒストグラムで保持する。ある弱識別器において、特徴量を x 、 x に応じたヒストグラムのビンを j 、Positive 及び Negative クラスの確率密度を W_+^j 、 W_-^j とすると、弱識別器の応答値は式 (5) の $h(x)$ となる。

$$h(x) = \frac{1}{2} \ln \frac{W_+^j + \varepsilon}{W_-^j + \varepsilon} \quad (5)$$

式 (5) において、 ε は十分に小さな係数であり、分母が 0 になることを抑止する。 $h(x)$ は、 W_+^j が W_-^j より大きければ正の符号をとり、反対に小さければ負の符号をとる。また、 $h(x)$ が正のとき、 W_+^j の値が W_-^j の値より大きいほど、 $h(x)$ は大きな値をとる。この応答値 $h(x)$ の大きさは、 x に対する弱識別器の出力の信頼度である。図 3 中の $h_1(x)$ の場合、 x の値がグラフの中央より左側であれば、 $h_1(x)$ は正の値をとる。また、 x がグラフの中央から左に向かうに連れて W_+^j は大きく W_-^j は小さくなり、 $h_1(x)$ の値は大きくなる。

二つ目の弱識別器の分離度 α は、ヒストグラムの全てのビンの確率密度から式 (6) により計算する。

$$\alpha = 1 - \sum_i \sqrt{W_+^i W_-^i} \quad (6)$$

式 (6) の α は 0 から 1 の値をとる。また、 W_+^i と W_-^i の確率密度関数の重なりが小さいほど大きな値をとる。この分離度 α は、弱識別器の確率密度関数の信頼度である。図 3 を例にとると、Positive と Negative の両クラスの確率密度関数の重なりが小さな $h_1(x)$ 、 $h_2(x)$ 、 $h_T(x)$ の順に、 α の値は大きい。

以上で計算した応答値 $h(x)$ と分離度 α は、式 (7) を用いて弱識別器の信頼度 $e(x)$ に集約する。 $e(x)$ の値が大きな弱識別器は、そのフレームにおいて、高い信頼度で人領域内の人の部分を捉えている。

$$e(x) = \begin{cases} \alpha h(x) & \text{if } h(x) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

次に、小領域ごとの信頼度は、小領域内の弱識別器の信頼度から式 (8) の E で計算する。この小領域の分割は、3.1 で述べた人領域の検出と同一である。

$$E = \sum_{x \in S} e(x) \quad (8)$$

式 (8) 中の S は人領域内の小領域、 $x \in S$ は S 内の特徴量 x の集合を示す。 E の値は、小領域内に大きな信頼度の弱識別器が多数あるほどに大きくなる。

信頼度で荷重した CRIM は、以下のように計算する。まず信頼度を乗じた濃度こう配を式 (9) の G_E で計算する。

$$G_E = \begin{bmatrix} E_1 G_1^T & E_2 G_2^T & \dots & E_B G_B^T \end{bmatrix}_{n \times 3}^T \quad (9)$$

式 (9) において、 B は人領域内の小領域の総数、 E_k 、 G_k は k 番目の小領域の信頼度と時空間のこう配値のスタックである ($k = 1, 2, \dots, B$)。 G_E の例を図 4 に示す。 G_E を求めた後は、2.2 の手順で CRIM を計算する。この重み付け CRIM は、人領域の検出の信頼

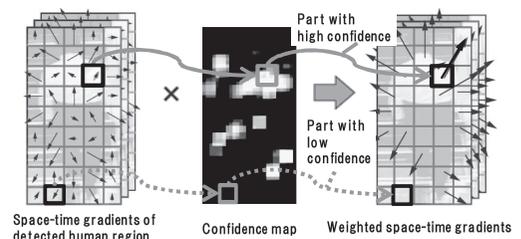


図 4 信頼度で重み付けた時空間のこう配値の例
Fig. 4 Example of space-time gradients weighted by confidence.

度を用いて人領域内の人の見えと動きに大きな重み値を与える。これにより、背景の動きの影響を低減する。

3.3 非定常度を用いた異常行動の検知

提案手法は各フレームで検出した全ての人領域の各々で CRIM を計算する。そして、その CRIM が一つでもしきい値を超えれば、異常行動として検知する。

4. 評価実験

提案手法の有効性を以下の実験で評価した。

4.1 評価実験の概要

評価実験の背景は実験設備のエスカレータとした。背景中では、定常的にステップが昇降する。異常行動は、乗降時の転倒とした。この転倒はエスカレータで代表的な異常挙動であり、救急事故の過半を占める [18]。転倒のデータは、6名の演技者が模擬で行った。

カメラは、このエスカレータの斜め後から撮影した。奥行によって、手前の人と奥側の人とで画像上の大きさが変化する。カメラは 360×240 [pixel] \times 30 [fps] の映像を取得し、これを評価実験の入力とする。

評価の対象は、提案手法二つと、従来手法二つとした。提案手法は、以下の二つとした。

- 提案手法 1: 信頼度の重み付き CRIM による手法
 - 提案手法 2: 信頼度の重みなし CRIM による手法
- 提案手法 1 と 2 で、人領域の検出は共通である。人領域の検出の後、提案手法 1 は 3.2 の方法で、提案手法 2 は 2.2 の方法で CRIM を計算する。CRIM を計算した後の異常行動の検知は、提案手法 1 と 2 共に 3.3 の方法で行う。各フレームで、人領域内の CRIM が一つでもしきい値を超過すれば異常行動を検知し、それ以外のときは異常行動を検知しない。

従来手法は、背景の動きの影響の抑止と異常行動の検知に有効な従来の技術を適用した以下の二つとした。

- 従来手法 1: 背景の除去有の CHLAC による手法
 - 従来手法 2: 背景の除去無の CHLAC による手法
- 従来手法 1 は、動的な背景から対象物を小領域単位で検出する背景差分法 [9] により対象物の領域を抽出した後に、残る背景を除去した映像を入力に用いる。従来手法 2 は、カメラの映像をそのまま入力とする。

各手法の詳細は、以下のとおりである。提案手法 1 と 2 の人領域の検出の学習には、Positive クラスを 2002 個、Negative クラスを 6048 個のサンプルを用いた。サンプルの内容は、3.1 の説明のとおりである。従来手法 1 と 2 は、CHLAC の累積時間は予備実験から文献 [3] と同じ 20 フレームと定めた。部分空間の

構築には、人が正常に通過する 3000 フレームを用いた。部分空間の次元は、累積寄与率 0.99 を条件とした。従来手法 1 では 5 次元、従来手法 2 では 14 次元となった。

4.2 代表ケースの評価結果

4.2.1 提案手法の代表ケースの評価結果

提案手法の代表ケースの評価結果の CRIM のグラフと代表フレームを図 5 に示す。グラフの傾向を述べると、提案手法 1 は異常行動のフレームでしきい値より大幅に高く、提案手法 2 は異常行動のフレームでしきい値をわずかに超える場合と超えない場合がある。

図 5 の 100, 308, 686, 982, 1267 の代表フレームは順に、通過する人、手前を横切る人、手前側で大きな動きで転倒した人、奥側で大きな動きで転倒した人、手前側で小さな動きで転倒した人である。前二つは正常行動、後三つは異常行動である。転倒を補足すると、大きな動きの転倒は、転倒した後に上体がステップの近くまで倒れたことで、姿勢の崩れが大幅である。小さな動きの転倒は、転倒のすぐ後に手をステップに突き体を支えたことで、姿勢の崩れが小幅である。この中で 686 フレームの転倒は、位置が手前側かつ動きが大きいことより、見かけの動きが大きい。686 フレームの転倒と比較すると、982 及び 1267 フレームの転倒の見かけの動きは小さい。それぞれ、位置が奥側及び転倒の動きが小さい。これら代表フレームにおいて、提案手法 1 は正常行動と異常行動を全て正しく判定した。一方、提案手法 2 は、982 と 1267 フレームの転倒が未検知であった。

ここで図 5 の代表フレームを対象に、信頼度が与える人の部分への重みを評価した。まず、人領域から人の部分の画素を手作業で抽出した。次に、人の部分と人領域全体の信頼度の総和を求めたところ、後者に対する前者の比は、図 5 の左から順に 71, 79, 62, 57, 66%, 平均で 67% であった。これらの比は、信頼度を使う提案手法 1 の CRIM の人の部分への重みである。一方、人の部分と人領域全体の画素数の比は、図 5 の左から順に 39, 59, 41, 31, 35%, 平均で 41% であった。これらの比は、信頼度を使わずに人領域から均一に求める提案手法 2 の CRIM の人の部分への重みである。最後に、信頼度を使うときと使わないときを比べると、前者がどのフレームでも優れ、平均では 26% 優れた。

4.2.2 従来手法の代表ケースの評価結果

まず、図 6 に CHLAC の入力のフレーム間差分の 2

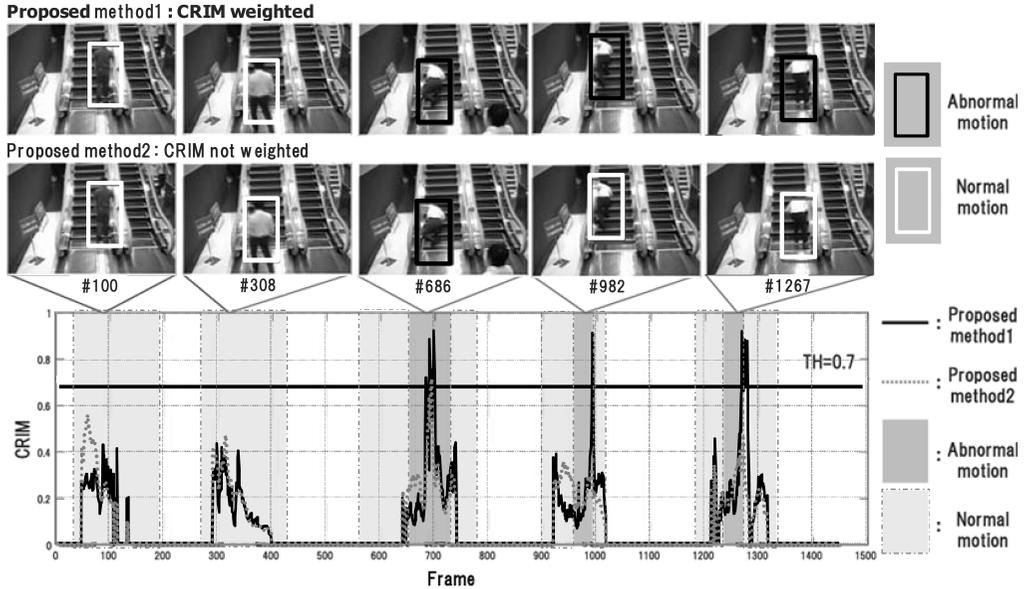


図5 提案手法の代表ケースの評価結果
Fig. 5 Result of example cases using the proposed methods.

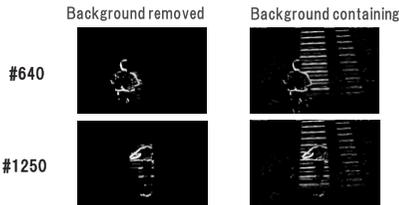


図6 2値化したフレーム間差分の例
Fig. 6 Examples of binarized frame subtraction.

値画像の例を示す。640フレームはステップに近づく途中、1250フレームは転倒後にステップに突く腕を伸ばした瞬間である。背景を除去した左側は、人の輪郭をおおよそで捉えている。一方、背景を含んだ右側は、人の輪郭に加えて、ステップを捉えている。他のフレームも同様の傾向にあった。

次に、図7に従来手法の評価結果の部分空間からの逸脱度のグラフと代表フレームを示す。代表の各フレームの行動は図5と同一である。逸脱度のグラフを見ると、従来手法1は正常行動と異常行動の過半を分離しているが、従来手法2は両者の分離が明瞭でない。代表フレームにおいて、従来手法1は282フレームの誤検知と982フレームの未検知があった。前者は、横切りの発生がカメラに近く、見かけの動きが大きいためと考える。後者は、反対にカメラから遠く、見かけ

の動きが小さいためと考える。従来手法2は282フレームが誤検知の上、転倒は三つどれも未検知である。

4.3 定量評価

定量評価は二つの動作を対象とした。転倒とかがみである。かがみは、図8のようにステップの前まで歩いて近づいた後に、直立から一息にかがむ動作である。転倒より見目の動きが大きい。動作の種類が与える影響を測るために評価の対象とした。

転倒の評価には、正常行動18ケース、異常行動12ケースを用いた。かがみの評価には、正常行動18ケース、異常行動9ケースを用いた。正常行動は、ステップを通過する動作である。通過する人数(1~4人)、通過する位置、ステップの方向(上り・下り)を変化させた。転倒は、通過する人数、ステップの方向、転倒の位置と動きの大きさを変化させた。かがみは、かがむ位置とステップの方向を変化させた。各手法で異常行動検知のしきい値を変えたときの検知率と誤検知率を図9のReceiver Operating Characteristicカーブ(以下ROCカーブ)に示す。ROCカーブでは、グラフが左上にあるほどに性能が優れる。図9(a)のかがみは手法間の差が小さい。一方、図9(b)の転倒は手法間で差が大きく、提案手法1が最も左上側にある。

詳細な比較を、表1に示す誤検知率10%を代表点にとった検知率を対象に行う。まずかがみは、差が小さ

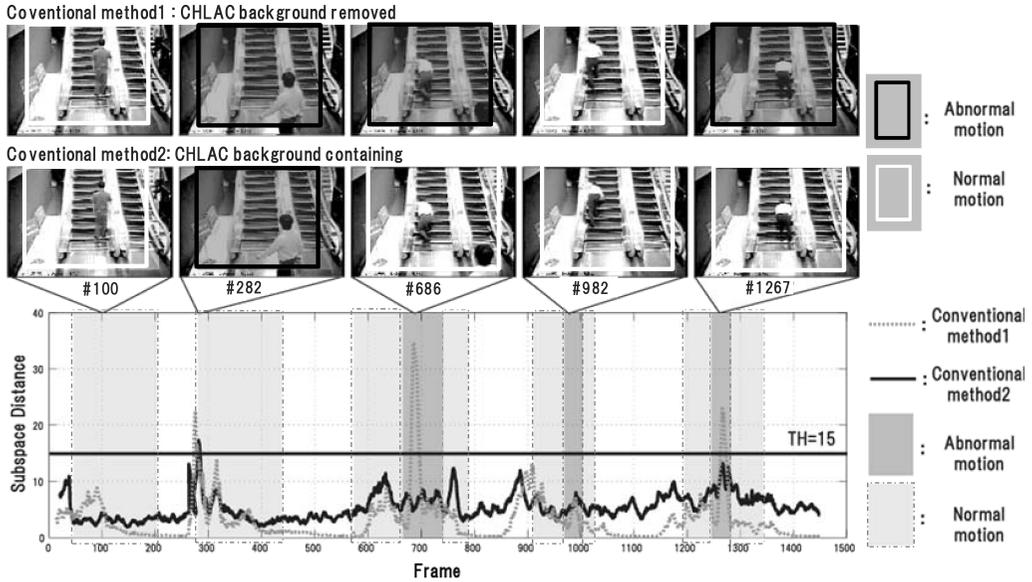


図 7 従来手法の代表ケースの評価結果
 Fig. 7 Result of example cases using the conventional methods.

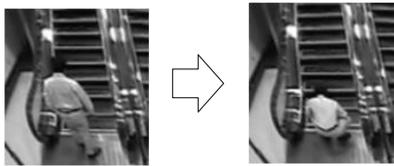


図 8 かがみの画像の例
 Fig. 8 Example image of crouch.

い。従来手法 2 以外は 89% という高い値、従来手法 2 も 78% という良好な値を得た。これは、見えが大きな異常行動は、背景の動きの影響を受けにくいことを示す。次に転倒は手法間の差が大きい。そのうちで、提案手法 1 が最も高い値 83% を得た。これは、従来手法 1 の 67% と比較して 16% 優れる。提案手法 1 と従来手法 1 のかがみと転倒の評価結果と、転倒の方がかがみよりも見かけの動きが小さいことを合わせると、前者は後者より見かけの動きが小さな異常行動の検知で優れた特質があると考えられる。提案手法 1 が提案手法 2 より 33% 優れることと、人領域の検出が両者で共通であることより、信頼度の重み付けがこの特質に関与すると考えられる。なお、提案手法と従来手法の差の要因は、後に 4.7 の評価結果を含めて述べる。

定量評価を通じて、検知が不良なケースの傾向は、提案手法 2 と従来手法 1 と 2 は 4.2 の代表ケースと同様であった。見かけの動きが小さな異常行動の未検知、

表 1 誤検知率 10% 時の異常行動の検知率
 Table 1 Detection rate of abnormal motion at false detection 10%.

Method	Proposed1	Proposed2	Conventional1	Conventional2
Feature	CRIM		CHLAC	
Suppression of dynamic background	Human region + weight	Human region	Background subtraction	Nothing
Crouch	89%	89%	89%	78%
Stumble	83%	50%	67%	42%

及び見かけの動きが大きな正常行動の誤検知が主だった。提案手法 1 には、通過の途中の不意の行動への誤検知があった。ステップの手前の振り返りが一例である。これは今後、異常行動を検知した後に動き特徴量等から行動認識をして、非定常度の高い正常行動と異常行動を区分すれば解決が可能と考える。また、検知した人領域の手前を、他の人が横切る瞬間の動きに誤検知があった。これは今後、複数の人領域間の遮へいを判定し、手前の人動きの影響を抑制すれば解決が可能と考える。なお、複数人が通過する途中で転倒するケースでも、提案手法は個別の人領域を検出して転倒を検知できた。ただし、これは通過する人数が少なく、転倒する人が手前の人にほぼ遮へいされなかったためと考える。人相互の遮へいが頻繁な状況、例えば混雑時へ適用することは、今後人領域の検出に上半身の身体の一部を捉える方法の適用で可能と考える。

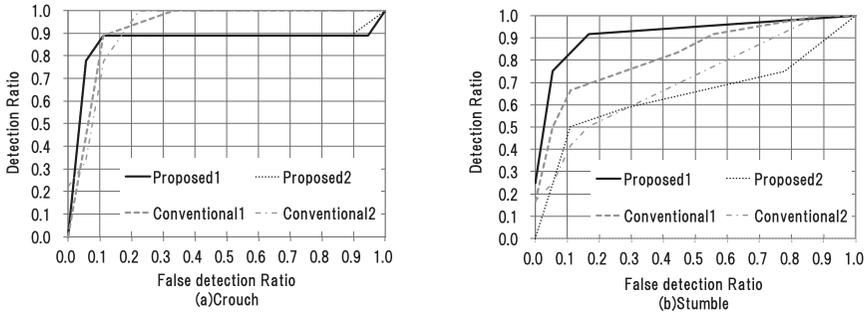


図9 ROC カーブ
Fig. 9 ROC Curves.

4.4 処理時間の評価

処理時間の平均値を 2.5 GHz の CPU 1 コアで計測した。提案手法は 1 と 2 が同一で 195 ms であった。内訳は、53 ms が ST-Patch 特徴量の計算、139 ms が人領域の検出、3 ms が CRIM の計算である。従来手法 1 と 2 は 48 ms と 23 ms であった。処理時間については、提案手法は従来手法より劣る。ただし、提案手法は処理領域ごとに並列処理が可能なので、複数コアの適用でリアルタイム化が可能である。6 コアを使えば、全画面を処理できる。なお、このコア数は処理領域を限定すれば削減できる。

4.5 人の大きさの影響の評価

人領域中の人の大きさが提案手法に与える影響を転倒の異常行動の検知率で評価した。まず、転倒開始時の人の足元の Y 座標を基準に評価ケースを手前と奥に同数のグループに 2 分割した (Y 座標 166 が境界)。人領域の平均の身長は、手前が 130 画素、奥が 110 画素で、前者は後者より 1.2 倍大きい。次に、提案手法 1 と 2 で、誤検出率 10% 時の奥と手前の検知率の内訳を表 2 のように求めた。表 2 において、まず提案手法 2 の手前と奥の検知率を比べると、両者で差がない。これは信頼度の重みのない非定常度は、人の大きさのばらつきをある程度許容することを示す。次に提案手法 1 と 2 の検知率を比べると、前者は後者よりも手前と奥の両方で優れる。これは信頼度の重み付けが、人の大きさのばらつきをある程度許容して有効であることを示す。最後に提案手法 1 の手前と奥を比べると、後者は前者より 17% 劣る。これは信頼度の重み付けが最大の効果を発揮するには、人の大きさが適切であることを条件とすることを示す。

4.6 人領域検出の評価

提案手法中の人領域の検出率と誤検出率を評価し

表 2 提案手法の転倒の検知率の内訳

Table 2 Breakdowns of stumble detection ratio of the proposed methods.

	Proposed1	Proposed2
Near side	50%	25%
Far side	33%	25%
Total	83%	50%

表 3 人領域の検出率

Table 3 Detection ratio of human regions.

		Near side		Far side	
		Normal behavior	Abnormal behavior	Normal behavior	Abnormal behavior
Stumble	Mean	96%	89%	89%	87%
	Minimum	64%	56%	53%	48%
Crouch	Mean	88%	80%	—	—
	Minimum	58%	4%	—	—

た。対象データは、4.3 の実験の延べ 6486 フレームとした。

まず、検出率での要求条件は、十分に高くして異常行動時における非定常度の変化を許容できることである。これは、図 5 のグラフのように非定常度が異常行動中に変化することと、非定常度がしきい値を超えた部分を逃すと異常行動が検知できないことによる。この要求条件が満たされないときは、異常行動の検知率は低下する。特に、検出率がゼロのときは、非定常度が一切抽出されないため、異常行動を全く検知できない。

表 3 に検出率の実験結果を示す。異常行動の種別 (転倒/かがみ)、人の大きさ (手前/奥、4.5 と同じ基準)、行動の種別 (正常行動/異常行動) の条件が影響することを想定して、条件ごとに平均値 (全体の傾向) と最低値 (最も不良な場合) を代表値で求めた。異常行動時は、動きが激しくて非定常度が高い、発生時点から姿勢を崩している間とした。なお、正常行動時の人領域検出は異常行動の検知に関与しないが、基本性

能として記した。人領域検出の成功は、画面中の人と人領域のずれが、画面中の平均的な頭部の大きさの 20×20 pixel 以下に収まることを条件とした。

表 3 中で、平均値はどの条件でも 80%以上と良好であった。最低値は転倒で 48%，かがみで 4%であった。転倒の 48%は、誤検知率 10%時に提案手法 1 が該当ケースで異常行動を検知できたので、今回の実験では前述の要求条件を満たせたと考える。一方、かがみの 4%は、提案手法 1 と 2 共に誤検知率が 100%近くでない該当ケースで異常行動を検知できなかったため、前述の要求条件を満たせなかったと考える。人領域の検出率が低い原因は、該当ケースでは実験者が腰を曲げた姿勢で歩いた後にかがんだことと、人検出の学習サンプルにこの姿勢が含まれていないことであった。今後、実システムの人領域検出の構築においては、カバーすべき人物の姿勢を十分に検討して、学習サンプルを収集する必要がある。

なお、かがみの最低値を除けば、異常行動時の検出率はどの項目でも、正常行動時から 2~8%低い程度であった。ここで、人領域検出の Positive の学習サンプルである直立に近い姿勢でステップを通過する人と、転倒並びにかがむときの人とを、人領域検出で用いる見えと動きの特徴で比較すると、後者の動きは身体が落下することで前者と大きく異なる。一方、後者の見えは、発生時点から大幅に姿勢を崩すまでの間は、前者から大きくは異なる。よって、提案手法中の人領域検出は、主に見えに基づくために、異常行動時でも大きくは性能が劣化しないと考えられる。

次に誤検出率は、誤検出した人領域が異常行動の偶発的な誤検知を頻発しないように、十分に低いことが要求条件となる。偶発的な誤検知とは、誤検出した人領域への人の進入や退出が偶発的に高い非定常度を招くケースで生じる。誤検出率の実験結果は転倒 7%、かがみ 11%であった。共に頻繁な異常行動の誤検知の誘発はなく、前述の要求条件を満たしていた。

4.7 動き特徴量の比較評価

提案手法中の CRIM と従来手法中の CHLAC を動き特徴量の単位で比較評価した。人領域内の CHLAC を用いる手法を比較手法 2、加えて CHLAC に人領域検出の信頼度を重み付けた手法を比較手法 1 とした。比較手法 1 及び 2 は、背景の動きの影響の抑止の方法が提案手法 1 及び 2 とそろう。比較手法 1 及び 2 の CHLAC では、累積時間を従来手法と同じ 20 フレーム、部分空間の構築用のデータを従来手法と同じ

表 4 誤検知率 10%時の異常行動の検知率の比較
Table 4 Comparison of detection rate of abnormal motion at false detection 10%.

Method	Proposed 1	Proposed 2	Compared 1	Compared 2
	CRIM		CHLAC	
Suppression of dynamic background	Human region + weight	Human region	Human region + weight	Human region
Stumble	83%	50%	75%	50%

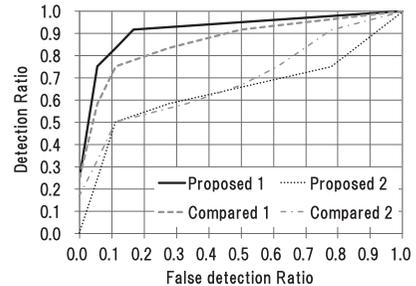


図 10 提案手法と比較手法の ROC カーブ
Fig. 10 ROC Curves of the proposed and compared methods.

人が正常に通過する 3000 フレーム、部分空間の次元を従来手法と同じ累積寄与率 99%の条件で 8 及び 15 とした。実験結果として、転倒の異常行動の誤検知率 10%時の検知率を表 4、ROC カーブを図 10 に示す。

提案手法 2 と比較手法 2 の検知率が、表 4 で共に 50%であることと、図 10 のグラフで誤検知率 10%から 50%の間ではほぼ一致することより、信頼度の重み付けがないと CRIM と CHLAC の差はほぼないと考える。一方で、提案手法 1 と比較手法 1 の検知率を比べると、前者は後者より表 4 で 8%優れ、図 10 のグラフで誤検知率 0%から 50%の間で優れた。両者の差には、CRIM が CHLAC よりも発生時間の短い異常行動の検知で優れることが影響していると考えられる。これは動き特徴量の累積時間のフレーム数が、CRIM が 5、CHLAC が 20 であり、前者の方が短くて発生時間の短い異常行動の検知に適すことによる。また、CRIM を用いた手法において信頼度の重み付けによって検知が可能となる見かけの動きが小さな転倒のうち、実際の動きの小さな転倒は発生時間が短いことによる。また、図 5 の 1267 フレームの動きの小さな転倒（転倒直後に手を突いて姿勢の崩れを軽減）の例では、誤検知率 10%時に提案手法 1 は検知できたが比較手法 1 は検知できなかったことによる（比較手法 1 は誤検知率 27%時に検知できた）。

加えて、比較手法 1 において CHLAC の累積時間を CRIM と同じ 5 フレームに短縮したときは、誤検

表 5 異常行動の検知率の平均値の比較
Table 5 Comparison of mean detection ratio of abnormal motion.

Method	Proposed	Compared	Conventional
Feature	CRIM	CHLAC	CHLAC
Suppression of dynamic background	Human region (+ weight)	Human region (+ weight)	Background subtraction (or nothing)
Stumble	67%	63%	54%

知率 10%時に図 5 の 1267 フレームの転倒を検知できた。ただし、CHLAC に増えたノイズの影響を抑止するためにしきい値を低くする必要があったので、全体の検知率が 50%と短縮前の 75%から大きく低下する悪影響を伴った。この結果からも、CRIM は CHLAC より発生時間の短い異常行動の検知で優れると考えられる。

最後に、表 1 と表 4 に示す誤検知率 10%時の提案手法、比較手法、従来手法の検知率の平均値を表 5 で比べた。提案手法が比較手法より 4%優れるのは、背景の動きの影響の抑止の方法が共通なので、動き特徴量の差である。比較手法が従来手法より 9%優れるのは、動き特徴量が共通なので、背景の動きの影響の抑止の方法の差である。また提案手法と従来手法の差 13%は、動き特徴量と背景の動きの抑止の方法の差の両方により、内訳は前者が 4%、後者が 9%である。

5. む す び

本論文では、見えと動きの情報を併せ持った ST-Patch 特徴量の統計的学習による人領域の検出と、統計的学習の信頼度で重み付けした非定常度を用いて、動的な背景下の突発的な異常行動を検知する手法を提案した。また評価実験より、提案手法の有効性を示した。今後の課題には、行動認識等による異常行動の検知の条件の拡充、交差や混雑で人領域同士が重なり合うケースへの対応による性能向上や適用シーンの拡充を挙げる。

謝辞 提案手法のアルゴリズムの研究に貢献頂いた株式会社サムスン横浜研究所（当時、株式会社日立製作所）の数井誠人氏、並びに評価映像の収集に協力頂いた株式会社日立ビルシステムの山口伸一朗氏に謹んで感謝の意を表する。

文 献

- [1] B. Scholkopf, J. Platt, J. Shawe-Taylor, A. Smola, and R. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Comput.*, vol.13, no.7, pp.1443–1471, 2001.
- [2] H. Zhong, J. Shi, and M. Visontai, "Detecting unusual activity in video," *Proc. Computer Vision and Pattern Recognition*, vol.2, pp.819–826, 2004.
- [3] 南里卓也, 大津展之, "複数人動画像からの異常動作検出," *コンピュータビジョンとイメージメディア*, vol.45, no.SIG_15, pp.43–50, 2005.
- [4] T. Kobayashi and N. Otsu, "Action and simultaneous multiple-person identification using cubic higher-order local auto-correlation," *Proc. International Conference on Pattern Recognition*, pp.741–744, 2004.
- [5] 関真規人, 林健太郎, 谷口博康, 橋本 学, 笹川耕一, "リアルタイム人物暴れ検出システム," *画像センシングシンポジウム*, 第 10 回, pp.E-02-1-6, 2004.
- [6] 都築勇司, 西村 孝, 藤吉弘亘, 森 次郎, "特徴点追跡結果を用いた動きのモデル化と例外行動検出," *画像センシングシンポジウム*, 第 10 回, no.IN1-07, pp.1–9, 2008.
- [7] E. Shechtman and M. Irani, "Space-time behavior-based correlation OR how to tell if two underlying motion fields are similar without computing them?," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.29, no.11, pp.2045–2056, 2007.
- [8] Y. Ke, R. Sukthankar, and M. Hervert, "Event detection in crowded videos," *Proc. International Conference on Computer Vision*, pp.8–15, 2007.
- [9] 弓場 竜, 三好雅則, 藤吉弘亘, 村井泰裕, "時空間テクニカを用いた背景モデルによる動体検知法," *信学論 (D)*, vol.J94-D, no.7, pp.1101–1112, July 2011.
- [10] M. Kazui, M. Miyoshi, and S. Muramatsu, "Incoherent motion detection using a time-series gram matrix feature," *International Conference on Pattern Recognition (ICPR)*, pp.1–5, Dec. 2008.
- [11] 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*, vol.1, no.2, pp.21–31, July 2008.
- [12] T. Furukawa, H. Fujihoshi, and A. Nomura, "A method for estimating cut-edit points in personal videos," *IEEE International Conference on Multimedia & Expo*, 2010.
- [13] Y. Murai and H. Fujiyoshi, "Shot boundary detection using co-occurrence of global motion in video stream," *International Conference on Pattern Recognition*, 2008.
- [14] 古川拓弥, 藤吉弘亘, "First Person Vision のための ST-patch 特徴を用いた自己動作識別," *パターン認識・メディア理解研究会*, pp.53–58, May 2010.
- [15] B. Wu, H. Ai, C. Huang, and S. Lao, "Improved boosting algorithms using confidence-rated predictions," *Mach. Learn.*, no.37, pp.297–336, 1999.
- [16] P. Viola and M. Jones, "Robust real-time face detection," *J. Comput. Vis.*, vol.57, no.2, pp.137–154, 2004.
- [17] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *Proc. Computer Vision and Pattern Recognition*, vol.1, pp.886–893, 2005.

- [18] 東京消防庁, “エスカレーターに係る事故防止対策について,” 2005.

(平成 23 年 8 月 31 日受付, 24 年 2 月 12 日再受付)



弓場 竜 (正員)

1997 京大・電機系学科卒. 1999 同大大学院工学研究科修士課程電子工学専攻了. 同年 (株) 日立製作所入社. 画像処理・画像認識に関する研究に従事.



村井 泰裕

2007 中部大・工・情報工学卒. 2009 同大大学院工学研究科情報工学専攻了. 同年 (株) 日立情報制御ソリューションズ入社. 画像情報処理・コンピュータビジョンに関する研究に従事. 情報処理学会会員.



藤吉 弘亘 (正員)

1997 中部大学大学院博士後期課程了. 1997~2000 カーネギーメロン大学ロボット工学研究所 Postdoctoral Fellow. 2000 中部大学講師を経て 2004 より同大学助教授. 2006 カーネギーメロン大学ロボット工学研究所客員研究員. 工博. 計算機視覚, 動画画像処理, パターン認識・理解の研究に従事. 2005 ロボカップ研究賞受賞. 2008 第 5 回日本 e-learning 大賞奨励賞受賞. 情報処理学会, 電気学会, IEEE 各会員.