平成26年度
中部大学大学院工学研究科情報工学専攻

博士学位論文

時空間特徴量と統計的学習を用いた動作認識に関する研究

弓場 竜

論　文　要　旨

本論文では，動画像中の見えと動きを捉える時空間特徴量と，背景や人の特徴を学習する統計的学習を用いることで，セキュリティ分野で課題となる環境条件の変化に頑健な動作像認識の手法を提案する．

監視カメラのデジタル化と共に動画像認識技術への要求がセキュリティシステムにおいて高まっている．この動画像認識技術では，現状広く実利用されている動体検知や滞留物検出に加えて，人等の対象カテゴリーの物体を検出する技術や動作を認識する技術等の動画像からより多くの情報を認識する先進的な技術が求められている．この様な先進的な動画像認識技術が実現できれば，カメラの映像を意味的に要約することにより，カメラの映像を目視で確認している監視者の負担の軽減や人為的ミスの回避に寄与できる．また，監視映像を対象とした動画像認識技術をマーケティング等のセキュリティ以外の用途に適用することで，セキュリティシステムの付加価値を高めることができる．

先進的な動画像認識技術を実現するには，環境条件の変化に対する 2 つの大きな課題がある．1 つ目の課題は，背景の見えの変化に対する頑健さである．背景の見えの変化に対する頑健さは，日照等の照明条件の変化により画像中の背景の明度が変化する時や，風に揺れる樹木等により画像中の背景が動く時等に求められる．2 つ目の課題は，人の見えの変化に対する頑健さである．人の見えの変化に対する頑健さは，カメラ視点の変化により画像中の人の向きが変わる時や，人の遠近により画像中の人の大きさが変化する時等に求められる．

これら課題に対し，本研究では動画像中の見えと動きを捉える時空間特徴量と統計的学習を用いた動作認識のフレームワークを提案する．また，動作認識のフレームワークを実現するために必要な動体検知，物体検出，および動作認識の各手法を提案する．まず，カメラ画像内の背景が大域的かつ局所的に複雑に変化する状況下において，時空間特徴量と統計的学習により背景の見えの変化をモデル化した背景差分による動体検知手法を提案する．これにより，背景が大域的かつ局所的に変化するときでも，誤検知を抑止して動体を頑健に検知することが可能になる．また，カメラ画像中を通過する人の見えと動きを時空間特徴の統計的学習により学習することで，背景が定常的に変化する場面において，人領域を検出した後に，人領域内の非定常度を評価することで転倒動作を検出する手法を提案する．これにより，背景中で定常的に動くエスカレータを誤検出することを抑止しながら，突発的に転倒する人の動作を検出することが可能になる．

さらに，本研究では，距離画像センサと呼ばれる画像中の各画素の距離値を計測するデバイスを適用し，距離情報を用いることで姿勢等による人の見えの変化に対して頑健な動作認識手法を提案する．提案手法では，距離画像を用いた動作認識において，学習時と認識時の間でカメラ視点が変化することで人の向きが変化することが及ぼす影響を生成的学習により抑止する．まず学習時のカメラ視点の距離画像の学習サンプルから，カメラ視点が変化した時の三次元データおよび時空間特徴量を生成することで，生成範囲内のカメラの視点に対応した弱識別器を構築する．そして，この弱識別器と認識時のカメラ視点における少数の学習サンプルから，認識時の人の向きに最適な強識別器を構築する．これにより，学習サンプルの収集の手間を小さく抑えながら，カメラ視点が変化した時の人の動作を良好に認識することが可能となる．加えて，本研究では距離画像センサを用いた

動作認識をより多くの場所に適用していくために，距離画像センサと対象とする人の距離が近い場合に，人の一部が画角外にはみ出て欠損する状況に対応可能な手法を提案する．提案手法では，欠損したときの時空間特徴量と全身が映るときの時空間特徴量の相関性を利用した回帰推定により時空間特徴量を補正することで動作認識の精度を改善する．これにより，エレベータ内の様に動作する人のすぐ近くにカメラが設置され，人の一部が画角外にはみ出やすい場面でも，良好に動作を認識することが可能となる．

目 次

図 目 次

表 目 次

第1章

序論

本章では，本論文の背景及び目的を述べた後，本論文の構成について述べる．

1.1　研究の背景

監視カメラのデジタル化と共に，セキュリティシステムにおける動画像認識技術への要求が高まっている．動画像認識技術では，現状広く実利用されている動体や滞留物の検出に加え，動画像からより詳細な情報を抽出する先進的な技術が求められている．このような先進的な技術の中に，人等の特定物体を検出する物体検出技術や人の動作を認識する動作認識技術が挙げられる．先進的な動画像認識技術が実現できれば，カメラ映像を意味的に要約することで，カメラ映像を目視で確認している監視者の負担軽減や人為的ミスの回避に寄与できる．そして，コンビニエンスストアや量販店等の流通分野や銀行等の金融分野，病院や鉄道駅等の公共施設や発電所や変電所等の重要施設といった多くの分野において，監視カメラを用いたセキュリティシステムの効率化を図ることが可能となる．また，監視映像を対象とした動画像認識技術をマーケティング等のセキュリティ以外の用途に適用することで，セキュリティシステムの付加価値を高めることができる．

動画像認識技術を現実のセキュリティシステムに適用するためには，環境条件の変化により生じる 2 つの大きな課題がある．1 つ目の課題は，背景の見えの変化に対する頑健さである．背景の見えの変化に対する頑健さは，日照強度等の照明条件の変化により画像中の背景の明度が変化する時や，風に揺れる樹木等により画像中の背景が動く時等に求められる．2 つ目の課題は，人の見えの変化に対する頑健さである．人の見えの変化に対する頑健さは，カメラ視点の変化により画像中の人の向きが変わる時や，人の遠近により画像中の人の大きさが変化する時等に求められる．

この環境条件の変化により生じる背景および人の見えの変化は，先進的な動画像認識技術ほど大きく影響を受けることになる．さらに，適用場所や認識対象に制約を受ける．動作認識技術の場合，現状広く適用されている範囲では，適用場所はエレベータ内の様な狭小なスペースに限られ，認識対象とできる動作の種類は動きが激しいために比較的に認識の容易な暴れに限られている．一方，動作認識技術では，より多くの監視カメラの映像を意味的に要約するために，適用場所の拡大が望まれている．また，より監視カメラの映像の要約度を高めるために，認識対象とする動作の種類の拡大が望まれている．セキュリティシステムにおいて，認識対象とする動作の種類は，事故や犯罪と結びついた種類の動作のニーズが最も高い．次いで，監視カメラから取得した認識結果をセキュリティシステム以外に応用するために，日常的な動作を含めたできる限り多くの種類の動作を認識対象とすることが望まれている．

1.2　研究目的

本研究では，時空間特徴量と統計的学習を用いた動作認識のフレームワークを提案し，フレームワークの構成要素である動体検知，物体検出および動作認識の各機能を実現することを目的とする．時空間特徴量とは，動画像中の 1 フレーム内の画像の見えに加えて，動画像のフレーム間で生じる動きの情報を捉えた特徴量である．この時空間特徴量を導入することにより，静止画像を対象とした手法で用いられる見えの情報に加えてフレーム間に生じた動きの情報を利用することが可能となり，動画像を効率的に認識することが可能になる．本研究ではこのフレームワークにより，環境条件

の変化により生じる背景および人の見えの変化に対して，頑健な動作認識技術の実現を目指す．そして，より広い範囲の場所を動作認識の適用場所とすることや，より多くの種類の動作を認識対象とすることで，セキュリティシステムにおけるニーズに応えることを目的とする．

本研究ではまず，カメラ画像中の背景が大域的かつ局所的に複雑に変化する状況下において，背景の変化の影響を抑止しながら背景差分により動体を検知する手法を提案する．セキュリティシステムにおいて監視カメラは様々な環境に設置されるため，動体検知では背景画像の多様な変化に対応できる手法が求められている．提案手法では，事前に背景中の変化を時空間特徴量を用いて学習しておくことで，背景の見えが変化しても誤検知を抑止しつつ動体を頑健に検知することが可能になる．

次に，本研究では背景中にエスカレータが定常的に動き続ける場面において，時空間特徴量の非定常度を利用して突発的に転倒する人の動作を検出する手法を提案する．セキュリティシステムにおいて，事故に直結するような人の動作は重要な監視対象であるため，この転倒等の動作を頑健に認識する手法には強いニーズがある．提案手法では，まず動画像中の小領域の中で時空間特徴量が人の条件を満たす人領域を検出する．人領域内における時空間特徴量を用いて定常的なのか非定常的なのかを判定することで，エスカレータのステップの動きと転倒の動作を判別する．

加えて本研究では，距離画像センサを用いることで，人の見えの変化に対して頑健な動作認識の手法を提案する．距離画像センサは画像中の各画素の距離を計測するデバイスであり，Time Of Flight(ToF) や Light Coding 等の幾つかの方式がある．2010 年に高精度かつ廉価な距離画像センサの製品が販売されたことで着目度が上がっている．距離画像センサにより取得する距離画像には，外光や影等の外乱の影響を受けにくい特長があると同時に，距離情報を用いて人の位置や形状を正確に捉えることにより動作認識を頑健にできるという特長がある．

距離画像を用いた動作認識において，本研究では学習時と認識時の間でカメラ視点が変化する場合に，人の向きの変化が及ぼす影響を生成型学習により抑止した動作認識の手法を提案する．ここでカメラ視点の変化は，動作する人とカメラの相対的な位置の変化から発生し，同一カメラ内において人の向きが変わる場合や，同一の人を異なる位置のカメラから撮影する場合に生じる．セキュリティシステムにおいて，前者は転倒や暴れといった事前に動作する人の向きが定まらない自発的な動作を認識対象とする場合に課題となり，後者は様々な箇所にカメラを設置する場合に課題となる．提案手法では，まず学習時のカメラ視点における距離画像の学習サンプルからカメラ視点を変化させた際の三次元データおよび時空間特徴量を生成することで，生成範囲内のカメラ視点に対応した弱識別器を構築する．そして，この弱識別器と認識時のカメラ視点における少数のサンプルから，認識時の人の向きに適した強識別器を構築する．これにより，学習サンプルの収集の手間を小さく抑えながら，カメラ視点が変化した時の人の動作を良好に認識する．

さらに，本研究では，距離画像センサを用いた動作認識を多くの場所に適用していくために，距離画像センサと動作する人の距離が近い場合において，人の一部が画角外にはみ出て欠損する状況に対応した手法を提案する．エレベータ内の監視カメラは動作する人の近くにカメラを設置するケースがあり，このようなセキュリティシステムでは人の身体の一部が画角外にはみ出すため，時空間特徴量も欠損することになり，動作認識が困難となる．そこで，提案手法では欠損したときの時空間特徴量と欠損しないときの時空間特徴量の相関性を利用した回帰推定により，時空間特徴量を補正す

ることで動作認識の精度を改善する．

1.3　本論文の構成

本論文は，図 1.1 に示す 7 つの章で構成されている．1 章では，本論文の背景と目的を述べた．2 章では，本論文で用いる時空間特徴量について述べる．以下，3 章と 4 章では，カメラ画像を対象とした動画像認識技術，5 章と 6 章では距離画像を対象とした動画像認識技術について述べる．3 章では，時空間特徴量を用いた背景モデルによる動体検知について述べる．4 章では，時空間特徴量と統計的学習を用いた異常動作検出について述べる．5 章では，距離画像の生成型学習を用いることでカメラの視点の変化に追従した動作認識について述べる．6 章では，人の身体の一部が欠損したときの時空間特徴量を回帰推定を用いて補正することで，欠損した状況に対して頑健な動作認識について述べる．7 章では，本論文の結論と展望について述べる．

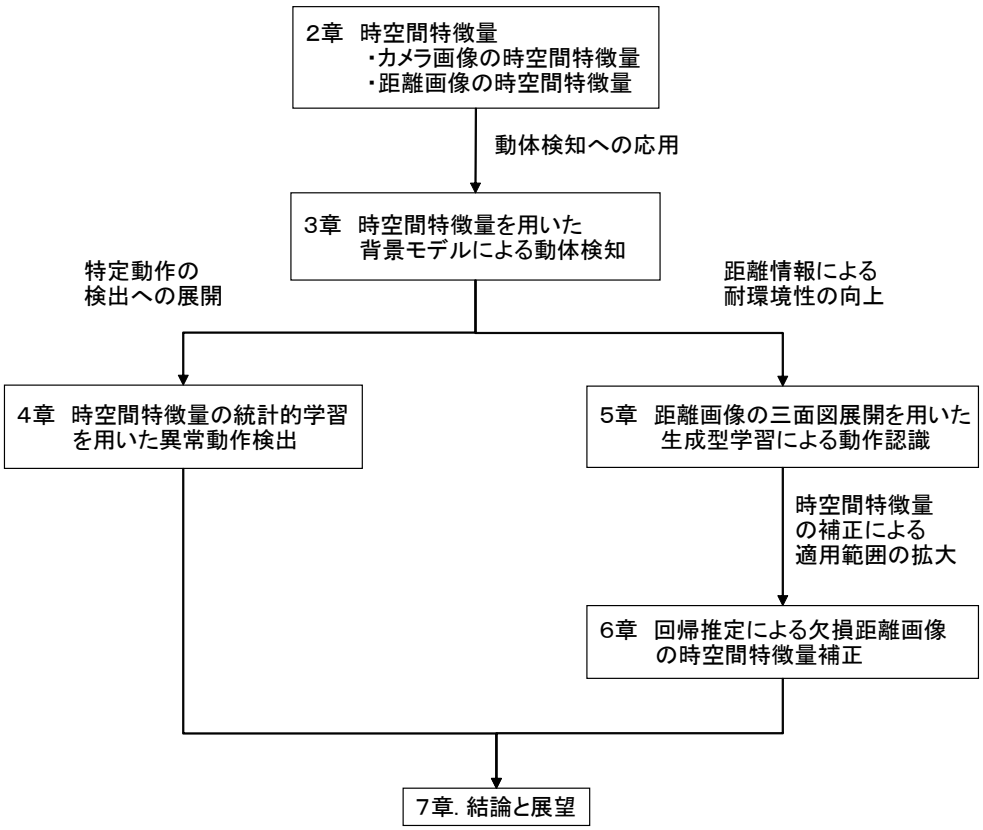


図 1.1: 本論文の構成.

第2章

時空間特徴量

時空間特徴量は，動画像中の１フレーム内の見えの情報とフレーム間の動きの情報を同時に捉える特徴量である．この時空間特徴量は，動体の進入や人の動作等の時間的に変化する対象を認識する際に適した特徴量である．時空間特徴量を用いた動画像認識は，カメラ画像を対象とした手法が主であったが，距離画像センサの出現を受けて，距離画像データを対象とした手法も増えてきている．

本章では，まずセキュリティシステムを主な適用先とした際の時空間特徴量と統計的学習を用いた動作認識のフレームワークを述べる．このフレームワークでは，動体検知，物体検出，動作認識の各機能を構成要素とする．次に，このフレームワークに適用可能なカメラ画像および距離画像を用いた従来の時空間特徴量のサーベイについて述べる．さらに，カメラ画像と距離画像の時空間特徴量の比較について述べる．

2.1 時空間特徴量の統計的学習を用いた動作認識のフレームワーク

セキュリティシステムを主な適用先とした際の時空間特徴量と統計的学習を用いた動作認識のフレームワークを図 2.1 に示す．本フレームワークには，大きく分けて学習フェーズと認識フェーズがある．学習フェーズでは，時間的に連続した複数のフレームのカメラ画像もしくは距離画像の学習サンプルから時空間特徴量を抽出し，抽出した時空間特徴量を入力とした統計的学習により動体検知，物体検出，動作認識の各識別モデルを構築しておく．認識フェーズでは，まず学習のフェーズと同様にカメラ画像もしくは距離画像から時空間特徴量を抽出する．次に，画面内に事前に設定した検知エリア内の時空間特徴量を動体進入と背景変動を判別する識別モデルに照らし合わせることで，検知エリアへの動体の進入を検知する (動体検知)．次に，動体が進入した検知エリア内の時空間特徴量を物体と背景を判別する識別モデルに照らし合わせることで，対象カテゴリの物体を検出する (物体検出)．最後に，検出した物体領域内の時空間特徴量を対象とし，動作の種類を判別する識別モデルに照らし合わせることで，物体の特定の動作を認識する (動作認識)．加えて，動体検知，物体検出，動作認識の情報を時間的に累積して関連付けることで，物体の行動を認識することができる (行動認識)．この行動認識は，動作認識と比べてより高度な動きの情報を認識することを指す．例えば，野球の試合を認識対象とする場合，まず球場内の所定箇所の動体を検知し (動体検知)，動体を検知した個所から選手やボールを検出し (物体検出)，各選手の投球や打球や走塁 (動作認識) の情報を収集すれば，野球の試合の状況を認識することができる (行動認識)．

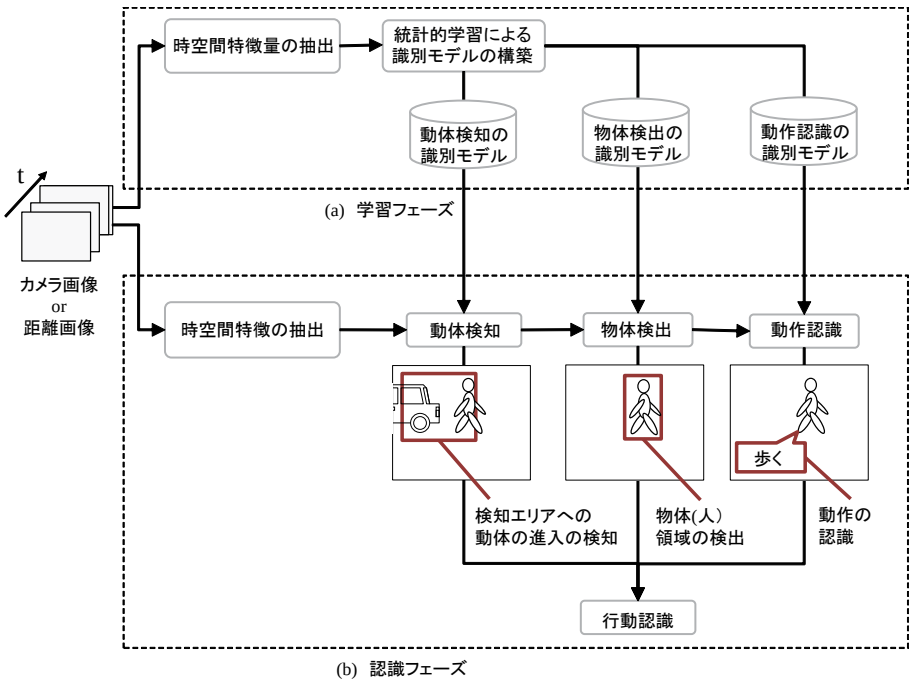


図 2.1: 時空間特徴量の統計的学習を用いた動作認識のフレームワーク.

本論文では，図 2.1 に示すフレームワークの中で，動体検知，物体検出，動作認識の範囲を研究対象とする．動体検知，物体検出および動作認識は，図 2.1 におけるフレームワークのサブセットであるが，それぞれ単独でセキュリティシステムにおいて有用となる技術である．動体検知は，監視エリアへの侵入を検知する用途，例えば発電所等の重要施設，留守時の家屋への侵入の検知において有用である．物体検出は，特定種類の物体を検出する用途，例えば立ち入り禁止箇所に進入した人や顔を検出する用途において有用である．動作認識は，犯罪や事故に直結する動作の認識，例えば暴力犯罪に結びつく殴り合いや，事故に結びつく転倒等の動作を認識する用途において有用である．

2.2 カメラ画像を対象とした時空間特徴量

本節では，時間的に連続したカメラ画像を対象とした時空間特徴量について述べる．カメラ画像を対象とした時空間特徴量は，表 2.1 に示すように，動画像中において時空間特徴量を抽出する空間的および時間的な単位の広さに応じて，特徴点，小領域，移動軌跡，ボリュームに大別できる．以下，この大別毎に代表的な時空間特徴量について述べる．

表 2.1: カメラ画像を対象とした時空間特徴量の種別．

種別	動画像中における領域	一例
特徴点	狭い	コーナー点，人体の関節位置
小領域	中	小パッチ
移動軌跡	中	移動ベクトル，追跡線
ボリューム	広い	時空間のボクセル

2.2.1 特徴点単位の時空間特徴量

本節では，時間的に連続したカメラ画像中の特徴点を単位とした時空間特徴量について述べる．特徴点を単位とした時空間特徴量は，動画像において所定の条件を満たす特徴点を抽出し，その特徴点もしくはその特徴点の周囲の動画像から時空間特徴量を抽出する．

■ Space-Time Interest Points

Space-Time Interest Points[1] は，ハリスのコーナー検出 [2] を時空間に拡張した手法であり，動画像中の特徴点を検出する手法である．Space-Time Interest Points を検出するには，フレーム t の

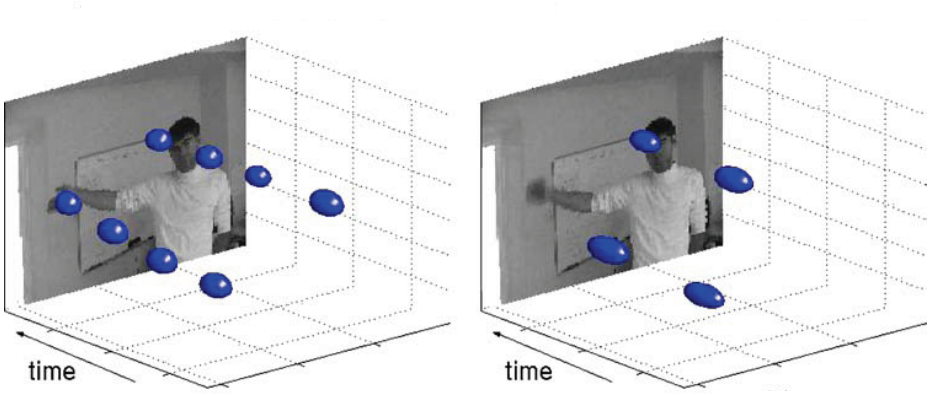
画像座標を (x, y) としたとき，行列 μ を式 (2.1) のように求める．

$$\mu = \begin{bmatrix} L_x^2 & L_x L_y & L_x L_t \\ L_x L_y & L_y^2 & L_y L_t \\ L_x L_t & L_y L_t & L_t^2 \end{bmatrix} \tag{2.1}$$

L は，ガウシアンフィルタをかけたシーケンスから算出する 1 次微分関数である．式 (2.2) に示すコーナー検出関数 H の極大値となる行列 μ の (x, y, t) を時空間における特徴点とする．

$$H = \det(\mu) - k \operatorname{trace}^2(\mu) \tag{2.2}$$

ここで，k は認識対象とするカメラ画像に応じて，適切なコーナ点が検出されるように実験的に調整する必要のあるパラメータである．図 2.2 に，Space-Time Interest Points の検出例を示す．図 2.2 に示す例では，同じ時空間の特徴点の分布から手を振る動作においても速く振る場合と遅く振る場合を識別することが可能であることを示している．



(a) 手を振る速度：速い　　(b) 手を振る速度：遅い

図 2.2: Space-Time Interest Points の検出例 (文献 [1] から引用)．

文献 [3] では，Space-Time Interest Points で検出した特徴点を中心とした時空間パッチから，Histogram of Oriented Gradients(HoG)[4] 特徴と呼ばれる勾配方向のヒストグラムと，Histogram of Optical Flow(HoF)[5] 特徴と呼ばれるオプティカルフローのヒストグラムを抽出し，一般物体認識等で用いられる Bag of Features(BoF)[6] と呼ばれる統計的学習を用いて映画シーン中の動作認識を実現する手法が提案されている．

2.2.2　小領域単位の時空間特徴量

本節では，時間的に連続したカメラ画像中の小領域を単位とした時空間特徴量について述べる．小

領域を単位とした時空間特徴量は，動画像中の小領域毎に，輝度値の勾配演算等により小領域内の画像内の見えや動きの特徴量を抽出する．

■ Space Time Patch 特徴量 (ST-Patch 特徴量)

Space Time Patch 特徴量 (ST-Patch 特徴量)[7] は，動画像の空間的および時間的な局所のパッチを単位としてパッチ内部の見えと動きを捉える特徴量である．ST-Patch 特徴は，画像を時間方向に重ねた時空間画像の局所領域における，各 x, y, t 軸の勾配値より求められる．画像中の各画素からは，図 2.3 に示すように，P_x, P_y, P_t の 3 つの勾配値を持つ時空間勾配が得られる．フレーム間で明度変化がない時，この時空間勾配は，各画素の移動方向と直交する．この時空間特徴をある局所領域の画素数でスタックした行列 G にその転置行列 G^T を掛けた行列を M とする．

$$\mathrm{M}=\mathrm{G}^T\mathrm{G}=\begin{bmatrix} \sum P_x^2 & \sum P_xP_y & \sum P_xP_t \\ \sum P_yP_x & \sum P_y^2 & \sum P_yP_t \\ \sum P_tP_x & \sum P_tP_y & \sum P_t^2 \end{bmatrix} \tag{2.3}$$

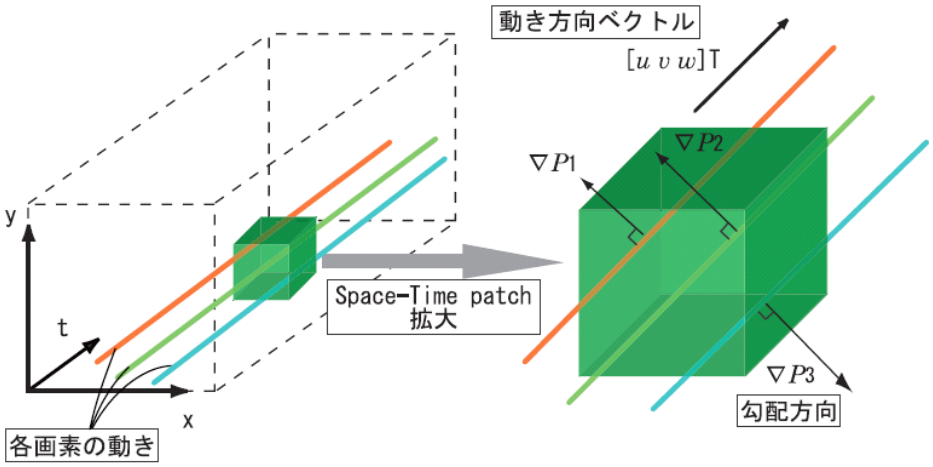


図 2.3: ST-Patch 特徴量．

行列 M が 1 つの ST-Patch から得られる特徴量となる．行列 M の左上の 2×2 行列で見えの情報，それ以外で動きの情報を表現している．ここで，行列 M は 3×3 の対称行列であるため，行列 M から重複しない行列要素を抽出すると式 (2.4) に示す 6 次元のベクトルとなる．

$$\boldsymbol{v} = \left[\sum P_x^2, \sum P_xP_y, \sum P_y^2, \sum P_xP_t, \sum P_yP_t, \sum P_t^2\right] \tag{2.4}$$

ST-Patch 特徴量を用いて所定の動作を検出するには，テンプレート動画の各部位の ST-Patch 特徴量と入力動画像の各パッチの類似度を用いて探索することにより実現できる．図 2.4 は，動画像中において人が手を振る，飛び跳ねる，歩く動作の検出結果の例を示している [7]．その他，ST-Patch 特徴量は，イベントの検出 [8]，動作の識別 [9]，移動方向別の物体検出 [10]，カット編集点の推定 [11]，シーン境界の検出 [12]，歩行状態の識別 [13] 等への適用が提案されている．

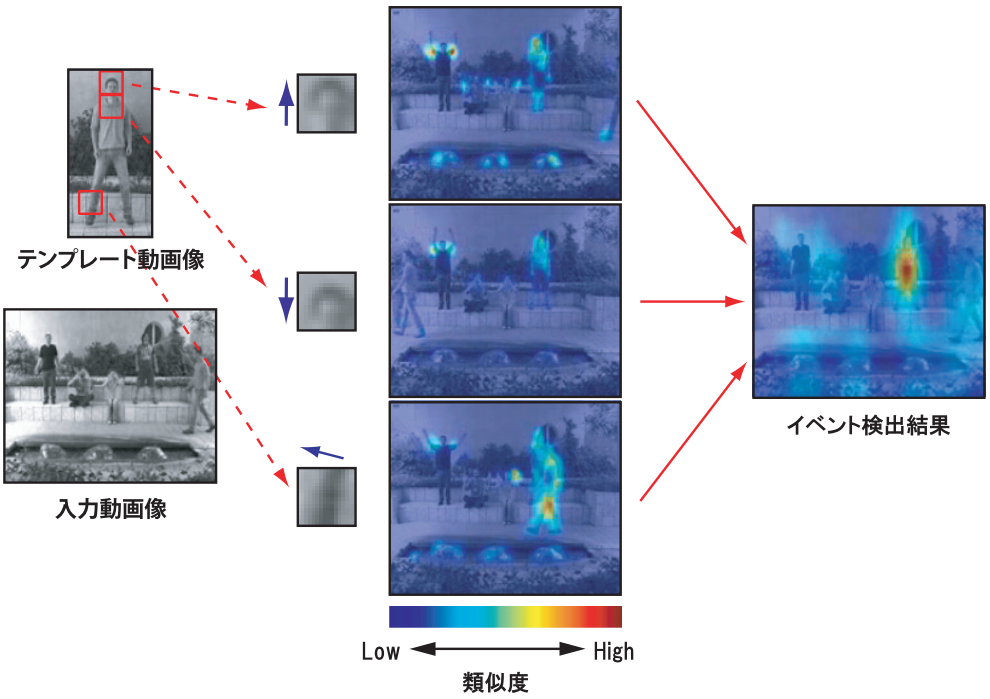


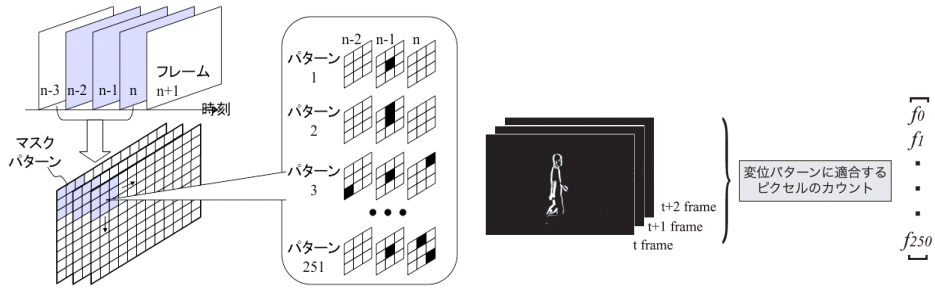
図 2.4: ST-Patch による特定動作の検出例 (文献 [7] から引用)．

■ Cubic Higher oder Local Auto-Correlation(CHLAC)

Cubic Higher oder Local Auto-Correlation(CHLAC) 特徴量 [14] は，静止画像内の局所領域内のテクスチャ情報を捉えた HLAC 特徴 [15] に時間軸を加えて三次元に拡張したものであり，動画像中の局所領域内の見えと動きを表現する時空間特徴量である．CHLAC は，立体高次局所自己相関とも呼ばれる．CHLAC 特徴の N 次自己相関関数の定式化の形は式 (2.5) のようになる．

$$x_f^N(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N) = \int f(\boldsymbol{r}) f(\boldsymbol{r} + \boldsymbol{\alpha}_1) \ldots f(\boldsymbol{r} + \boldsymbol{\alpha}_N) d\boldsymbol{r} \tag{2.5}$$

ここで，f は時系列 2 値化画像，$\boldsymbol{r}$ は画像内の 2 次元画像座標，N 個の変位パターン $\boldsymbol{\alpha}_i (i = 1, \ldots, N)$ は動画像の三次元座標 (x, y, t) の局所 3 近傍のベクトルを示す．図 2.5(a) に CHLAC の変位パターンの例を示す．CHLAC 特徴量は，時間的に連続した 3 つのフレーム間差分画像から，図 2.5(b) に示すように変位パターン毎に適合するピクセルをカウントして求める．



(a) CHLAC特徴量の変位パターン　　(b) CHLAC特徴量の抽出

図 2.5: CHLAC 特徴量.

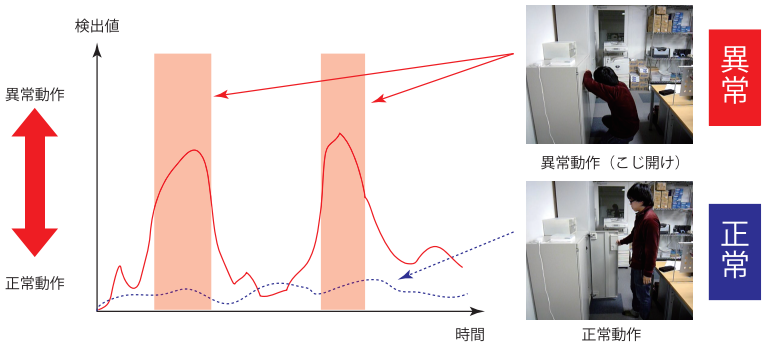


図 2.6: CHLAC による異常動作の検知例.

文献 [16] では，正常動作から抽出した CHLAC 特徴から構築した部分空間を用いて，盗難等の異常動作検知を実現している．異常動作のサンプルは頻繁に観測されないため学習に利用することが難しい．そこで，学習過程では，通常動作から求めた CHLAC 特徴のみを用いて，部分空間を予め構築しておく．判定時には，入力動画像の CHLAC 特徴を作成した部分空間に射影し，距離を求める．これにより，図 2.6 に示す様に，距離としての検出値が大きい場合を異常動作として検知することが可能となる．文献 [17] では，この CHLAC を応用したエレベータ内の異常動作検知が提案されている．

2.2.3　移動軌跡単位の時空間特徴量

本節では，時間的に連続したカメラ画像中の移動軌跡を単位とした時空間特徴量について述べる．移動軌跡を単位とした時空間特徴量は，画像中において複数フレーム間に渡り追跡した対応点の移動軌跡の情報を利用する．

■ フローベクトル

フローベクトルは，画像中において隣接した 2 時刻間の対応点を探索することで，画像中の移動の向きと方向を計測する手法である [18]．動体検知の基本的な方法として，監視カメラの記録フレームの制御などに広く用いられている．また，文献 [19] では，エレベータのかご内のフローベクトルを算出し，そのフローベクトルの個数，方向，強度の分散が正常の範囲を逸脱することを条件に，暴れの異常動作を検知する手法が提案されている．

■ 密な移動軌跡の特徴量

密な移動軌跡を用いる手法では，図 2.7(a) に示すように，各フレームにおいて複数解像度ピラミッド化した画像中から格子状の密な特徴点を抽出してから，これを複数フレームに渡って追跡することで密な移動軌跡の特徴量を抽出する．そして，各移動軌跡の近傍から HoG，HoF, および Motion Boundary Histogram(MBH)[20] と呼ばれる動きの境界を捉える特徴量を用いて，密な移動軌跡の特徴量 (dense trajectory description) を抽出する [21]．文献 [21] ではこの密な移動軌跡と BoF[6] と呼ばれる統計的学習法を用いることで，映画中の動作 [22] や web 上のビデオクリップの日常的な動作 [23] を認識している．

また，文献 [24] では，Difference of Gaussian(DoG) と呼ばれるスケールスペースの極値検出法を用いて画面中から特徴点を抽出し，特徴点それぞれを SIFT 特徴量 [25] と Mean Shift 法 [26] を用いて複数フレームに渡って追跡することで，画面中の主に人がいる領域から密な移動軌跡を取得し，この移動軌跡の密度と移動方向を特徴量として抽出する方法が提案されている．文献 [27] では，この移動軌跡の特徴量を所定時間集計することで，人流の密度と方向のビジュアライゼーション，ならびに正常時のフローベクトルの方向を統計的学習することによってモデルを逸脱する異常動作の検知に用いる方法が提案されている．

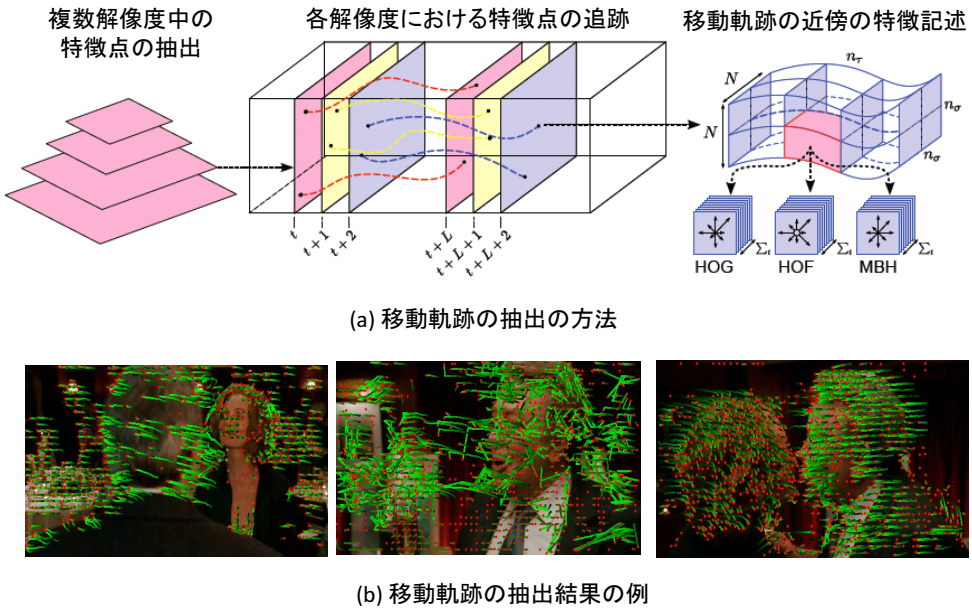


図 2.7: 密な移動軌跡の抽出の方法および抽出結果の例 (文献 [21] から引用).

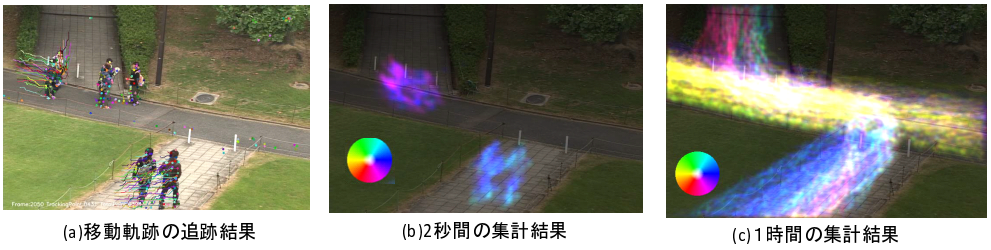


図 2.8: 特徴点の移動軌跡の追跡結果と集計結果の例 (文献 [27] から引用).

2.2.4 ボリューム単位の時空間特徴量

本節では，時間的に連続したカメラ画像中のボリュームを単位とした特徴量について述べる．ボリュームを単位とした時空間特徴量では．画面内の広範囲の領域を複数フレームに跨り累積したボリュームから見えと動きの特徴量を抽出する．

■ **Motion History Image(MHI)**

Motion History Image(MHI)[28] は，画像上の変化の発生時刻の履歴を濃淡値で記録した特徴量である．ある時刻 t の座標 (x, y) における MHI の値 $H(x, y, t)$ は，毎フレーム式 (2.6) の様に更新する．式 (2.6) において，$D(x, y, t) \in \{0, 1\}$ は，時刻 t の座標 (x, y) における変化の有無を示し，$D(x, y, t) = 1$ のとき変化が発生し，$D(x, y, t) = 0$ では変化は発生していない．ここで，$D(x, y, t)$ での変化とは，2 値化したフレーム間差分が一例である．また式 (2.6) において，τ は変化のタイムアウト時間である．式 (2.6) により MHI では， 現フレームから τ フレーム以内の範囲において，古い時刻で変化が発生した画素の明度が低く，手前の時刻で変化が発生した画素の明度が高く表現される．また，現フレームから τ フレーム以内に変化が生じなかった画素の明度は 0 になる．

$$H(x, y, t) = \begin{cases} \tau & \text{if } D(x, y, t) = 1 \\ \max(0, H(x, y, t-1) - 1) & \text{if } D(x, y, t) = 0 \end{cases} \tag{2.6}$$

MHI では，式 (2.7) を用いて現フレームからタイムアウト時間 τ 内に 1 回以上変化が生じた画素を抽出し，画像をセグメンテーションすることができる．この MHI を用いてセグメンテーションされた領域は，Motion Energy Image(MEI) と呼ばれる．

$$E(x, y, t) = \begin{cases} 1 & \text{if } H(x, y, t) > 0 \\ 0 & \text{if } H(x, y, t) = 0 \end{cases} \tag{2.7}$$

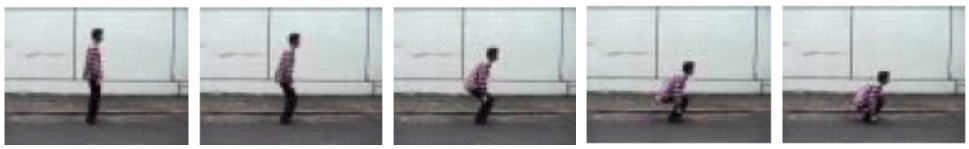
人が屈む動きをする時のカメラ画像と MHI と MEI の例を図 2.9 に示す．図 2.9 の (b)MHI では屈む運動を通じて動き続ける頭部や脚部の明度が高く，(c)MEI では屈む動作を通じて人の動きが生じた画素が抽出されていることがわかる．

以上求めた MHI および MEI からは，Hu Moment 法 [29] で 7 次元のモーメントを計算することで時空間特徴量を計算する．この Hu Moment 法には，平行移動および面積に対して不変という特長がある．

また MHI からは，画像中の勾配値を利用することで，動きの方向の特徴を抽出できる．ここでは，まず MHI 画像の各画素 (x, y) において，縦方向および横方向の濃度勾配 $F_x(x, y), F_y(x, y)$ を求め，次にこれら 2 つの濃度勾配の成す角度 $\phi(x, y)$ を式 (2.8) で求める．この勾配方向 $\phi(x, y)$ は，各画素における動きの方向を示す．画像中において，領域単位の特徴量を抽出するには，領域内の各画素における勾配方向 $\phi(x, y)$ のヒストグラムを求める．このヒストグラムは，その領域内における大域的な動きの方向性を示している．

$$\phi(x, y) = \arctan(F_y(x, y)/F_x(x, y)) \tag{2.8}$$

文献 [28] では，この MHI を用いて歩行や屈むや手を振るといった基礎的な動作や，ジェスチャの動作を認識する手法が提案されている．



(a) カメラ画像

(b) MHI (Motion History Image)

(c) MEI (Motion Energy Image)

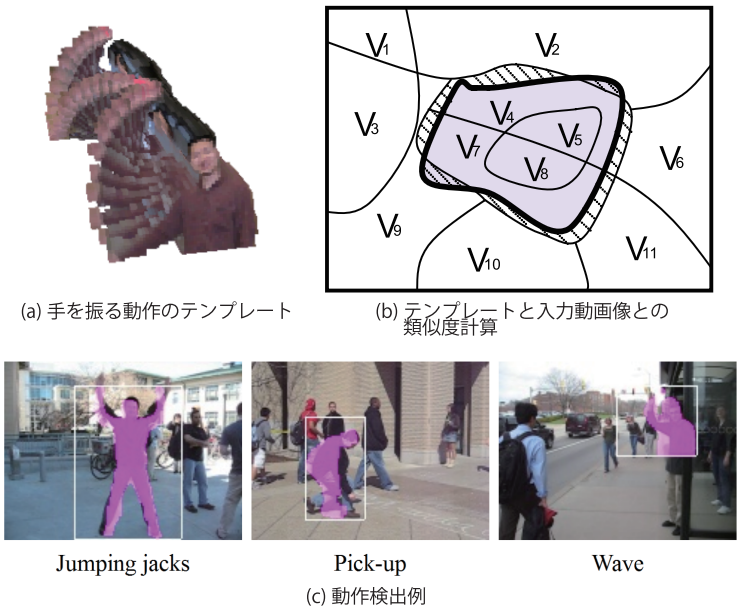
図 2.9: 人が屈む動きをする時のカメラ画像と MHI と MEI の例 (文献 [28] から引用).

■ Volumetric Features

文献 [8] では，セグメンテーションされた動画像とテンプレートを Volumetric Features と呼ばれる時空間特徴量を用いてマッチングすることで特定の動作を検出する手法が提案されている．まず，動画像を Spatio-Temporal Segmentation[30] を用いて動画像中の動きと形状情報を基にボリュームにセグメンテーションする．次に，セグメントされた動画像とテンプレートのボリュームデータをマッチングする．マッチングの際には，形状情報のマッチング結果と動き情報のマッチング結果を統合する．1 ステップ目では，形状情報のマッチングのために動画像のある位置とテンプレートとのマッチング距離を算出する．2 ステップ目では，動きの情報のマッチングにはセグメントされたボリュームにおいて前述の ST-Patch 特徴量を算出し，2 つのボリューム間の ST-Patch 特徴量の類似度をマッチング距離とする．これにより，テンプレートに登録された動作パターンと一連のボリュームデータとのマッチングにより動作を検出することが可能となる．

しかし，人の動作パターンは複雑であるため，多くの動きが含まれた動作パターンのマッチングは困難となる．そこで，ボリュームデータを任意の位置で細分化し，Parts-based Model のテンプレートを作成することにより，マッチング精度を向上させている．この手法は背景が複雑に動く環境下に

おいても動作検出が可能である．そのため，人通りが多い場所に設置されたカメラによる監視や支援システムへの応用が期待される．Volumetric Feature のテンプレート，類似度計算，動作の認識の例を図 2.10 に示す．



(a) 手を振る動作のテンプレート　(b) テンプレートと入力動画像との類似度計算

(c) 動作検出例

図 2.10: Volumetric Features の概要 (文献 [8] から引用)．

2.3　距離画像を対象とした時空間特徴量

本節では，距離画像を取得するための距離画像センサ，距離画像を対象とした時空間特徴量の大別，大別毎の時空間特徴量の代表例およびその適用先について述べる．なお，距離画像を対象とした時空間特徴量の応用先は，動作認識が主である．これは，距離画像を使う場合には，高精度かつ光学的な外乱に頑健な距離データが取得できるため，図 2.1 のフレームワーク中で動作認識の前段にある動体検知および物体検出は，多くの適用先において背景差分等の簡易な方法で実現できるためである．

2.3.1　距離画像センサ

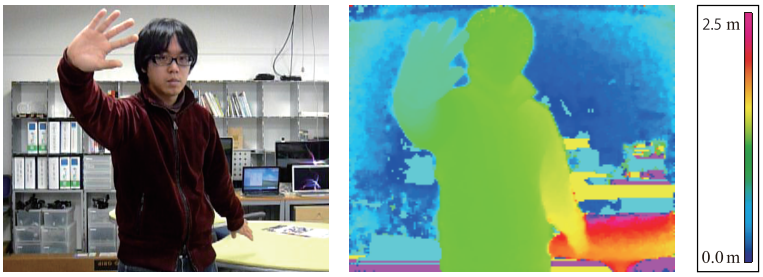
距離画像センサは特殊な光学系を用いて画像中の各画素の距離を計測するデバイスであり，距離画像はこの距離画像センサで取得された画像である．距離画像センサには幾つかの計測の方式が存在するが，以下に述べる 2 つの方式が代表的である．1 つ目の方式は，カメラから赤外光を照射し物体に反射してカメラに到達するまでの飛行時間を計測する Time of Flight(ToF) 方式 [31] である．2 つ目の方式は，特殊な赤外光によるパターンを照射し，そのパターンを三角測量により計測する Light

Coding 方式 [32] である．ToF 方式および Light Coding 方式の代表的なデバイスを図 2.11 に示す．また，Light Coding 方式で取得した距離画像の例を図 2.12 に示す．この距離画像センサの距離情報からは，従来のカメラの映像では困難であった三次元空間中の人の位置や形状等の特徴を捉えることができる．また，2010 年頃からこれら距離画像センサの廉価化が進み，セキュリティシステムをはじめとした実システムへの適用が期待されている．

(a) Swiss Ranger社 SR-4000
(ToF方式)

(b) Microsoft社 Kinect™
(LightCoding方式)

図 2.11: 距離画像センサの代表例．



(a) カラー画像　　(b) 距離画像

図 2.12: 距離画像の例．

距離画像には，さらに以下の 3 つの特長がある．

- 近赤外を自発光する光学系のために，低照度環境下において，カメラ画像で問題となってきたゲイン変動や撮像系のノイズの影響を受けない．
- 距離情報を扱うことで，影等の光学的な外乱の影響を受けない．
- 顔や服装の様なプライバシー情報が映らないために，公共な場所でのプライバシーの問題を避けやすい．

2.3.2　距離画像を対象とした時空間特徴量の種別

距離画像を対象とした時空間特徴量の大別を表 2.2 に示す．種別は，距離値，スケルトン，両者の複合に分けられる．距離値とは，距離画像中の各画素における距離画像センサから物体までの距離を示す．

種別のスケルトンからは，2.1 節のフレームワーク中の動作認識に強く関わる人の姿勢の情報を取得できる．更に，スケルトン周辺の距離値のデータを用いることで，より情報量の多い動作認識が可能となる．その一方，現状ではスケルトンの認識が容易にできる距離画像センサおよび計算機は限られるために，距離データのみを使う時空間特徴量と比べると汎用性の面で劣る．

表 2.2: 距離画像を対象とした時空間特徴量の大別．

種別	情報量	汎用性
距離データ	少	高
スケルトン	中	低
スケルトン＋距離データ	多	低

2.3.3　距離値を用いた時空間特徴量

距離値を用いた時空間特徴量では，距離画像中の各画素の距離値，ならびに空間的または時間的な近傍の画素の距離値を入力として，勾配値の計算等により人の見えと動きを抽出する．また，距離画像を用いた時空間特徴量では，距離画像中の各画素の距離値を，射影変換のモデルを用いて三次元座標 (x, y, z) に変換してから前述の演算を施すことで，三次元情報に基づいた人の見えと動きを抽出する．抽出した時空間特徴量は，統計的学習と組み合わせることで，人のジェスチャ等の意図的な動作や，歩行などの自発的な動作の認識に用いられる．以下，距離画像を用いた時空特徴量およびその適用例について述べる．

文献 [33] では，距離画像のフレーム間差分中の差分画素の三次元分布を用いた時空間特徴量が提案されている．この時空間特徴量は，Edit Distance 法による動作カテゴリ判別と組み合わせることで，ジェスチャの動作認識に用いられている．

文献 [34] では，距離画像中の人のシルエットを直上，正面，真横の 3 つの平面に投影した輪郭の形状を用いた時空間特徴量が提案されている．この時空間特徴量は，動作状態の遷移モデルを用いた動作カテゴリ判別と組み合わせることで，屈む等の基礎的な動作の認識に用いられている．

文献 [35] では，距離画像の小領域内の距離の最頻値を用いた時空間特徴量が提案されている．この時空間特徴量は，Joint-Boosting と呼ばれる識別器と組み合わせることで，商品棚から物品を取りだす動作の認識に用いられている．

文献 [36] では，距離画像中のキーポイントの座標値や距離値や移動方向等を BoF でベクトル量子化した時空間特徴量及び MHI のモーメントを用いた時空間特徴量が提案されている．これら時空間

特徴量は，Support Vector Machine(SVM) による動作カテゴリの判別と組み合わせることで，掃除等の日常動作の認識に用いられている．

文献 [37] では，距離画像の背景差分で求めたシルエット中のコーナー点の空間座標の時空間特徴量が提案されている (図 2.13) ．この時空間特徴量は，多様体による状態遷移モデルに依るカテゴリ判別と組み合わせることで，手を振る等の基礎的な動作の認識に用いられている．

文献 [38] では，距離画像中の人の表面から求めた法線ベクトルの時空間セル内における密度から求めたヒストグラムを用いた時空間特徴量が提案されている (図 2.14)．ここで時空間セルでは，動作カテゴリ間の判別が良くなる様に，粗密が最適化されている．この時空間特徴量は，SVM による動作カテゴリ判別と組み合わせることで，手のジェスチャ動作や，物を拾うや座る等の日常動作の認識に用いられている．

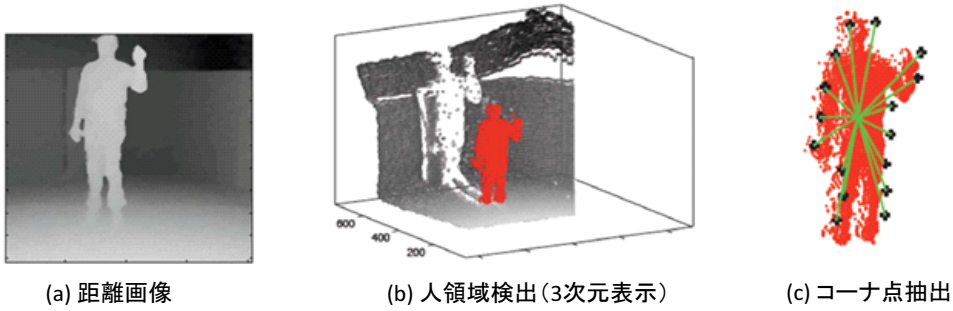


図 2.13: 距離画像中の人のコーナ点を用いた時空間特徴量 (文献 [37] から引用)．

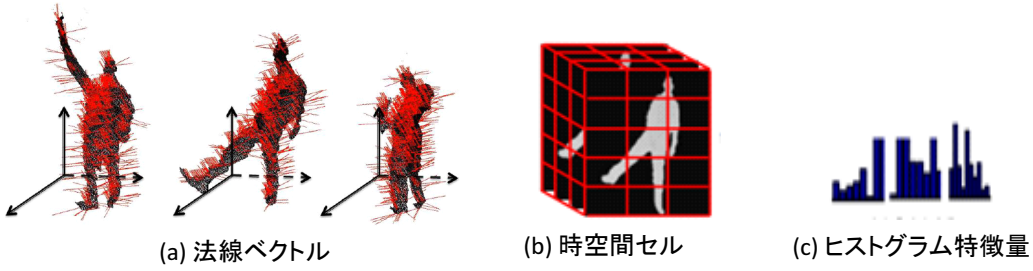


図 2.14: 距離画像中の人の法線ベクトルを用いた時空間特徴量 (文献 [38] から引用)．

2.3.4　スケルトンを用いた時空間特徴量

本節では，スケルトンの認識方法，スケルトンのみを用いた時空間特徴量，スケルトンと距離値を用いた時空間特徴量について述べる．

■ スケルトンの認識方法

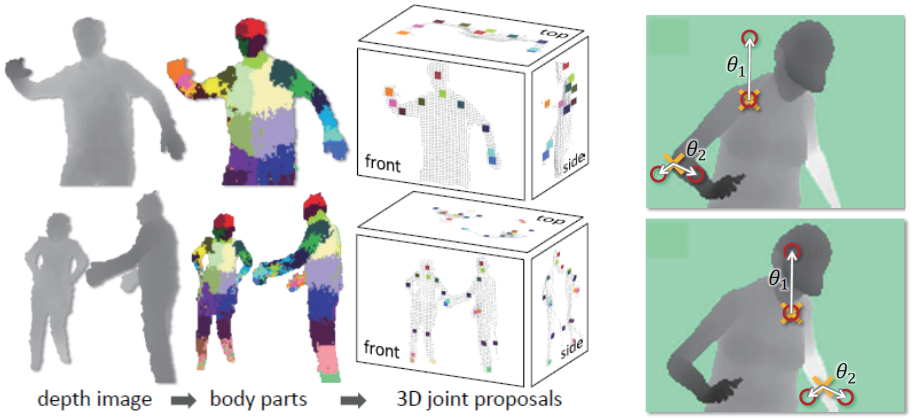
人のスケルトンとは，距離画像中の人それぞれの頭部や腕や足といった関節の位置を抽出するものである．スケルトン中の関節の位置からは，人の詳細な姿勢の情報が抽出できる．2011 年に，高精度かつリアルタイム性が高い手法が，文献 [39] にて発表された．また，文献 [39] のスケルトンの認識を実装した Software Development Kit(SDK) が Microsoft 社から提供されたことから，多くの研究で利用されている．この SDK は汎用の Personal Computer(PC) で信号処理可能な一方，同社の距離画像センサの Kinect$^{\mathrm{TM}}$ を用いることと，同社のオペレーションシステム上をプラットフォームとする制約を受ける．

文献 [39] では，学習の段階において，まず人の距離画像データの学習サンプルを人体モデルのコンピュータグラフィックから作成する．次に，学習用の距離画像から，特徴量として注目画素における任意の 2 画素の差を式 (2.9) により求め，決定木 (Random Forest) 中の各分岐ノードの条件とする．

$$f_\theta(I, \boldsymbol{x}) = d_I\left(\boldsymbol{x} + \frac{\boldsymbol{u}}{d_I(\boldsymbol{x})}\right) - d_I\left(\boldsymbol{x} + \frac{\boldsymbol{v}}{d_I(\boldsymbol{x})}\right) \tag{2.9}$$

I は画像番号，$\boldsymbol{x}$ は注目画素，$\boldsymbol{u}, \boldsymbol{v}$ は参照画素を指定する画像内のベクトル，$d_I(.)$ は各画素の距離値である．認識時には，まず，学習した決定木により身体パーツを画素毎に識別する，次に，画素毎の識別結果から Mean Shift クラスタリングによりパーツ位置を統合することで，身体パーツの各関節位置を推定する．姿勢推定の処理過程を図 2.15(a)，使用する距離特徴を図 2.15(b) に示す．

図 2.15(a) で示したスケルトンは，実際の距離画像の人からも高精度に抽出することが可能である．この手法で認識されたスケルトンは，ジェスチャ動作を用いたゲームのユーザインタフェースとして広く利用されている．



(a) 姿勢推定の処理過程　　(b) 距離画像からの特徴量

図 2.15: 距離画像中のスケルトンの認識 (文献 [39] から引用)．

■ スケルトンのみを用いた時空間特徴量

文献 [40] では，近接フレーム内の関節間の移動量を用いた時空間特徴量が提案されている (図 2.16)．この時空間特徴量は，図 2.16 に示す様な動作カテゴリ毎の代表フレームに対して高いスコアを出すように係数が求められたロジスティック回帰と組み合わせることで，屈む，走る，歩く等の基礎的な動作の認識に用いられている．

文献 [41] では，前段でスケルトンの所定の関節から人の重心と向きを推定し，鉛直および人の正面で向きを正規化した極座標 ($\phi - \theta$) を格子分割したヒストグラム中における関節の位置の出現頻度を用いた時空間特徴量が提案されている (図 2.17)．この時空間特徴量は，k-means および Hidden Markov Model(HMM) を用いたカテゴリ判別と組み合わせることで，歩行，座る，物を拾うといった基礎的な動作の認識に用いられている．

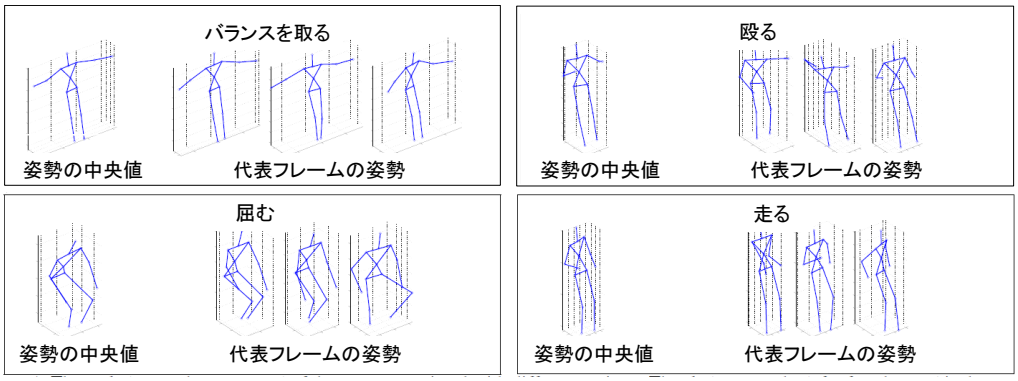


図 2.16: 各動作における代表フレームのスケルトン (文献 [40] から引用)．

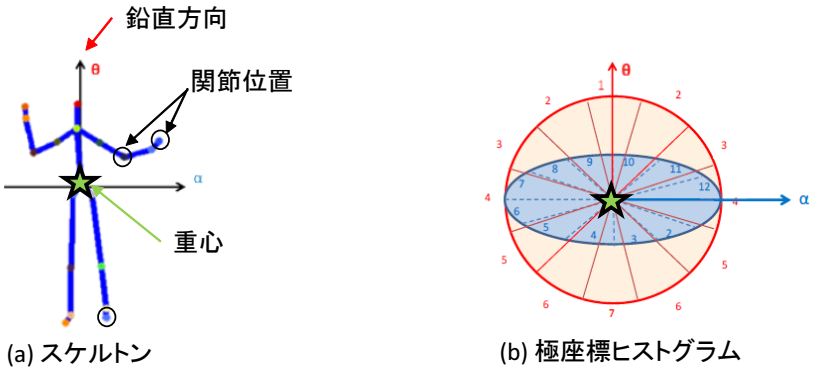


図 2.17: スケルトンに設けた極座標を用いた時空間特徴量 (文献 [41] から引用)．

■ スケルトンと距離値を用いた時空間特徴量

スケルトンと距離値を用いた時空間特徴量は，スケルトン中の関節の位置，およびスケルトンの周囲の距離値のデータから，人および人の所持する物品の見えと動きを捉えた特徴量を抽出している．この人と物品の見えと動きを捉えた時空間特徴量は，物品を持った人の動作を，人だけから特徴量を抽出するよりも良好に認識することが可能となる．

文献 [42] では，スケルトン中の関節の移動量と各関節の近傍の距離値のデータの密度を記述した Local Occupancy Patterns(LOP) を所定時間累積し，ピラミッド状に分割した時間窓内の短時間フーリエ変換の記述子 (Fourier Temporal Pyramid) に変換した時空間特徴量が提案されている (図 2.18)．この時空間特徴量は，Actionlet と呼ばれる関節のサブセットの選択法と Multiple Kernel Learning(MKL)[43] によるカテゴリ判別と組み合わせて，楽器の演奏等の物品が関与した動作の認識に用いられている．

文献 [44] では，距離画像に加えてカメラ画像を入力として，上半身のスケルトンの関節の位置の時系列の特徴量と，物体検出で検出した物体の位置や SIFT 特徴量 [25] 等による形状の特徴量を併せた時空間特徴量が提案されている．この時空間特徴量は，Markov random field(MRF)[45] を用いた状態遷移モデルと組み合わせることで，人の動作と物体および物体と物体の相互作用，およびこれら相互作用の時間変化を用いた時空間特徴量が提案されている (図 2.19)．この時空間特徴量は，構造化 SVM によるカテゴリ判別と組み合わせることで，食事を食べたり薬を飲むといった物品の関与する日常動作の認識に用いられている．

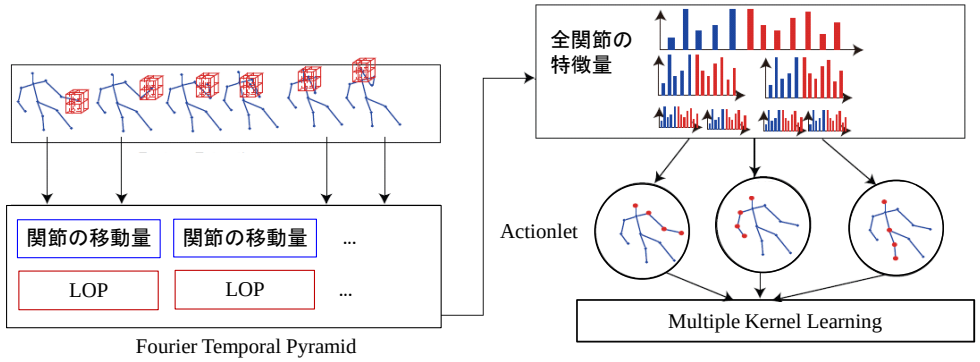


図 2.18: スケルトンと関節付近の距離値のデータ (LOP) を用いた手法の例 (文献 [42] から引用)．

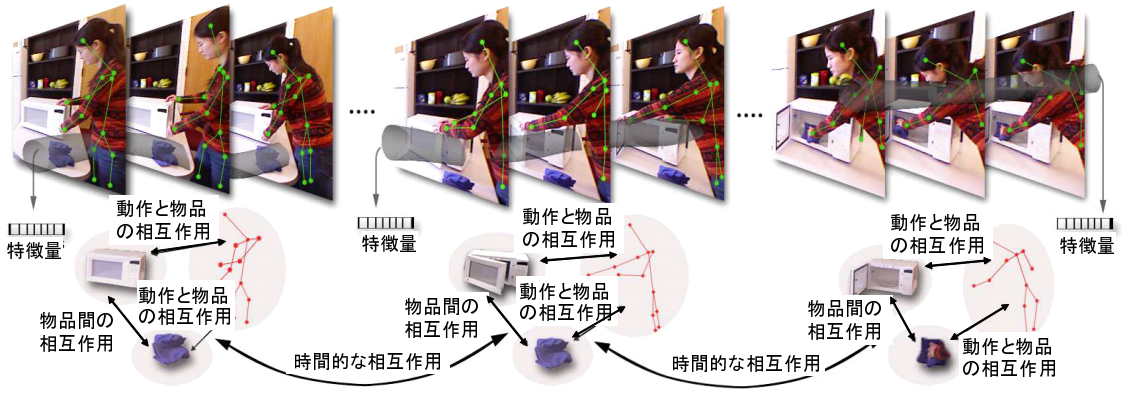


図 2.19: スケルトンと周囲の物品の相互作用の認識の例 (文献 [44] から引用).

2.4 カメラ画像と距離画像を用いた時空間特徴量の比較

本節では，カメラ画像および距離画像を用いた時空間特徴量の比較について述べる．図 2.1 に示した時空間特徴量を用いた動画像認識のフレームワークをセキュリティシステムに適用する場合，環境条件に対する頑健さ，および適用範囲の広さが重要となる．2.2 節および 2.3 節におけるサーベイの結果より，カメラ画像および距離画像の時空間特徴量の得失の比較を表 2.3 に示す．比較項目は，環境条件に対する頑健さでは背景と人の見えの変化を中項目として設け，各中項目から小項目を洗い出した．なお，本論文において，距離画像センサは，セキュリティシステムに適用することが比較的に容易な ToF 方式もしくは Light Coding 方式を想定した．表 2.3 において○は適用可能，△は条件付きで適用可能，× は適用困難を示す．

環境条件の変化に対する頑健さにおいて，距離画像を全ての小項目で○としたのは，近赤光を用いるため距離情報は影響を受けない，もしくは距離情報を用いて補正が可能であるからである．光学的な外乱の許容と，低コントラスト環境の許容の小項目が前者であり，残りの小項目が後者に該当する．反対に，環境条件の変化に対する頑健さにおいて，カメラ画像が全ての項目で△もしくは × となった項目は，カメラの画像の情報だけでは部分的にしか許容できない (△)，もしくは全く影響を許容できない (×) からである．

一方，どの小項目においても，適用範囲の広さが距離画像ですべて × である．これは，本論文で扱う距離画像センサは，近赤外光を用いた特殊光学系で距離値を取得することが原因である．この原因のために，視野角を広くしたり，距離計測の範囲を遠方まで広げたり，強い近赤外光を含む日光の照射を許容したり，筐体を小型化して僅かなスペースに設置することは困難である．

以上の比較結果をまとめると，カメラ画像の時空間特徴量は適用範囲の広さに優れるが，環境条件の変化への頑健さには劣る．一方，距離画像の時空間特徴量は，環境条件の変化への頑健さに優れるが，適用範囲の広さで劣る．この比較結果より，本論文ではカメラ画像および距離画像を用いた認識技術を次のように位置付ける．カメラ画像を用いる場合には，広い範囲のセキュリティシステ

ムに適用できる利点に着目し，対象とする認識機能を絞り，環境条件の変化に対して頑健な認識技術の実現を目指す．一方，距離画像を用いる場合には，適用範囲を絞り，環境条件の変化に頑健かつ機能面で優れた認識技術の実現を目指す．

表 2.3: 距離画像とカメラ画像の時空間特徴量の比較．

大項目	中項目	小項目	距離画像	カメラ画像
環境条件の変化への頑健さ	背景の見えの変化への頑健さ	光学的な外乱の許容	○	×
		低コントラスト環境の許容	○	△
		動揺する背景の許容	○	△
		周囲を通行する人の許容	○	△
	人の見えの変化への頑健さ	人の遠近の影響の許容	○	×
		周囲を通行する人の許容	○	△
		至近距離の人の見えの変化の許容	○	×
		カメラ視点の変化の許容	○	△
		人の向きの変化の許容	○	△
適用範囲の広さ	–	視野角の広さ	×	○
		遠方のカバー	×	○
		日光が照射する箇所への設置	×	○
		僅かな隙間への設置	×	○

2.5 まとめ

本章では，まず時空間特徴量と統計的学習を用いた動作認識のフレームワークについて述べた．このフレームワークでは，動体検知，物体検出，動作認識の認識機能を構成要素とし，セキュリティシステムを主な適用先とする．次に，このフレームワークに適用可能なカメラ画像および距離画像を用いた時空間特徴量について述べた．次に，カメラ画像と距離画像の時空間特徴量の比較について述べた．カメラ画像の時空間特徴量は適用範囲の広さに優れる一方で，環境条件の変化への頑健さで劣る．一方，距離画像の時空間特徴量は，環境条件の変化への頑健さに優れる一方で，適用範囲の広さで劣る．

本論文では，以上述べたカメラ画像および距離画像の時空間特徴量を用いて，動作認識のフレームワーク中の各認識機能を実現する手法について提案する．3 章と 4 章では，カメラ画像を用いて，対象とする認識機能を絞り，環境条件の変化の中で背景の見えの変化に対して頑健な手法を提案する．3 章では，背景変動に頑健な動体検知の手法を提案する．4 章では，背景が定常的に変化するエスカレータのシーンで，人領域を検出し，転倒という限られた種類の動作を認識する手法を提案する．5 章と 6 章では，距離画像を用いて，環境条件の変化に伴う人の見えの変化に対して頑健かつ多

くの種類の動作を認識できる動作認識の手法を提案する．5 章では，カメラ視点の変化により現れる人の向きの変化が，動作認識に与える影響を抑止する手法を提案する．6 章では，距離画像センサと動作する人の距離が近い場合に対応するために，人の一部が画角外にはみ出て欠損する影響を抑止する手法を提案する．

第3章

時空間特徴量を用いた背景モデルによる動体検知

動体検知とは，画像内に設けた検知エリア内に動体が進入することを検知する手法である．本章では，時空間特徴量を用いた背景モデルを用いることで，カメラ画像中の背景が大域的かつ局所的に複雑に変化する状況に有効な動体検知の手法について述べる．

背景差分は，動体検知の手法としてセキュリティシステムに広く適用されてきた．しかし，背景に大きな変化が発生した際に，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．そこで本章では，動画像中の小領域内の見えと動きの情報を反映した ST-Patch 特徴量と，混合正規分布を用いた背景モデルを生成することにより，背景の変化への追従性を高めた動体検知手法を提案する．評価実験では，屋外で日照が変化し樹木が揺れるシーン，屋内の天井照明が頻繁に明滅するシーン，屋外に面したエスカレータから乗客の身体が乗り出すシーンの 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することを確認した．また，提案手法を構成する構成要素の有効性と，構成要素が有効性を発揮する条件を明らかにした．

3.1 従来手法における課題

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への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背景モデルを生成する手法がいくつか提案されてきた [46][47][48][49][50][51][52]．これらの背景モデルの生成には，大別すると 2 つのアプローチがある．背景モデルの生成の 1 つ目のアプローチは，画像の最小単位である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手法である [46][47]．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等の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という利点がある．一方，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手法である [48][49]．また，小領域の単位の特徴量から頑健な背景モデルを構築するアプローチとして，小領域内の特徴量を複数状態の統計分布でモデル化する方法が提案されている [50][51]．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という利点がある．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 [52]．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 通りの手法で信号処理することが冗長で，背景モデルが複雑となり調整すべきパラメータが増加する課題がある．

ここで背景モデルの頑健さを高める 1 つの手法として，背景モデルの時間軸への拡張が提案されている．文献 [53] は，文献 [46] の画素の明度の確率分布モデルを時空間の隣接画素の明度との同時確率に拡張することで，信号雑音や局所的な背景の変化への頑健さを高めている．文献 [54] は，部分空間による小領域内の明度変化のモデルを時空間に拡張することで，夜間の照明の変化への頑健さを高めている．文献 [55] は，小領域内の特徴量の計算に時間予測した画素の明度を適用することで，急峻な日照の変化への頑健さを高めている．ただし，これら従来手法では，時間軸への拡張が部分的なために，限られた種類の背景の変化しか対応できていない．

3.2 提案手法のアプローチ

本章では，小領域内の見えと動きの情報を併せ持つ時空間特徴量の ST-Patch 特徴量と，複数状態を持つ混合正規分布を用いて頑健な背景モデルを構築することで，様々な種類の背景の変化に対して頑健な動体検知の手法を提案する．提案手法における背景モデルは，小領域内の見えの情報を用いて背景画像の大域的な変化に追従すると同時に，従来手法の背景モデルでは吸収が困難であった，背景画像の局所的な変化を動きの情報で吸収するアプローチをとる．図 3.1 に提案手法の流れを示す．

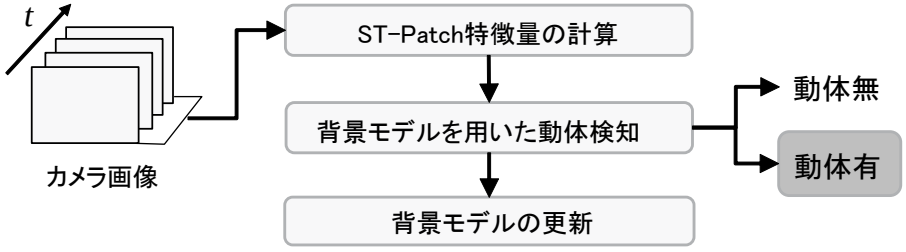


図 3.1: 提案手法の流れ．

以下，3.3 節では，提案手法の動体検知の手法について述べる．3.4 節では，提案手法の評価実験の実験条件について述べる．3.5 節では，提案手法の評価実験の実験結果について述べる．3.6 節では，本章のまとめを述べる．

3.3 ST-Patch 特徴量を用いた背景モデルによる動体検知

本節では，ST-Patch 特徴量を用いた動体検知の手法について述べる．以下，ST-Patch 特徴量の動体検知における特性，混合正規分布による背景モデルを用いた動体検知，背景モデルの更新について述べる．

3.3.1 ST-Patch 特徴量の動体検知における特性

本節では，ST-Patch 特徴量の動体検知における特性について述べる．ST-Patch 特徴量は，2.2.2 節で述べた様に，時間的に連続したカメラ画像中の小領域中の時空間の濃度の勾配値から見えと動きを捉えた特徴量である．提案手法では動体および背景の見えと動きを捉える時空間特徴量として，式 (2.4) に示した ST-Patch 特徴量の 6 次元の要素を用いる．

提案手法において，ST-Patch 特徴量は，見えの要素を用いて種類の異なる物体を区分する傾向を持つ．また，動きの要素を用いて，静止している物体と変化している物体の区分と，変化している

物体の間で変化の仕方が異なるものを区分する傾向を持つ．表 3.1 に ST-Patch 特徴量の例として，図 3.2 に示す樹木の同じ部分の (a) 静止した時と (b) 風に揺れる時，及び図 3.2 に示す路面の同じ部分の (c) 日照が安定している時と (d) 日照が変化している時の数値例を示す．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 3.1 において，ST-Patch 特徴量 $\{v_j\}_{j=1\sim6}$ には，下記の傾向が認められる．

- 時間変化の有無によらず，樹木および路面で見えの要素は分離．
  ((a) と (b) および (c) と (d) の $v_{1,2,3}$ の比較)
- 時間変化があると動きの要素が増大．
  ((a) と (b) の $v_{4,5,6}$，(c) と (d) の v_6 の比較)
- 日照の変化と樹の揺れとで動きの要素が相異．
  ((b) と (d) の $v_{4,5,6}$ の比較)

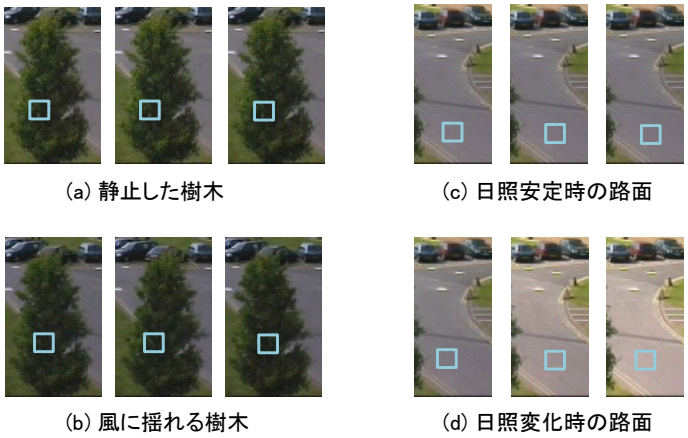


図 3.2: 背景の小領域の例．

表 3.1: ST-Patch 特徴量の数値例．

	見えの要素			動きの要素		
	v_1	v_2	v_3	v_4	v_5	v_6
	$\sum P_x^2$	$\sum P_x P_y$	$\sum P_y^2$	$\sum P_x P_t$	$\sum P_y P_t$	$\sum P_t^2$
(a) 静止した樹木	68.3	1.1	38.4	-0.2	0.1	0.6
(b) 風に揺れる樹木	66.5	-11.6	52.7	1.8	-1.3	1.7
(c) 日照安定時の路面	0.7	-0.1	0.6	0.0	0.0	0.0
(d) 日照変化時の路面	1.4	-0.1	0.7	-0.4	0.0	1.2

3.3.2 混合正規分布による背景モデルを用いた動体検知

本節では，まず，事前に収集したカメラ画像から，背景の変化の状態が複数ある場合を想定して，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方法について述べる．次に，入力画像中の 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について述べる．

■ ST-Patch 特徴量と混合正規分布による背景モデルの生成

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の枝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．そこで提案手法では，背景の変化が複数かつ有限個内の状態に収まることを前提条件において，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 を式 (3.1) および式 (3.2) に示す．

$$p(x) = \sum_{i=1}^{M} w_i p_i(x) \tag{3.1}$$

$$p_i(x) = \frac{1}{\sqrt{2\pi\sigma_i}} \exp\left\{ -\frac{1}{2}(x-\mu_i)^T \sigma_i^{-1} (x-\mu_i) \right\} \tag{3.2}$$

式 (3.1) および式 (3.2) において $p(x)$ は確率変数 x が混合正規分布に属する確率密度，$p_i(x)$ は確率変数 x が i 番目の正規分布に属する確率密度，w_i，μ_i，σ_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．提案手法では，ST-Patch 特徴量の各要素 $\{v_i\}_{i=1\sim6}$ 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，μ_i，σ_i は，EM アルゴリズム [56] 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．なお，計算した各パラメータが背景モデルとして適切であるためには，背景のサンプル映像の時間幅に，背景の変化のバリエーションの大半を包含する十分な長さが必要である．以上述べた ST-Patch 特徴量の混合正規分布のパラメータは，背景の変化が画像上の場所ごとに異なるために，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．

■ ST-Patch 特徴量による背景モデルを用いた動体検知

提案手法では入力映像が与えられた時，動画像の各フレームから検知エリア内の ST-Patch 特徴量を抽出し，背景モデルの混合正規分布から乖離したブロックを前景として検知する．提案手法が時空間の小領域内の特徴量を捉えるため，この前景には現フレームの動体および直前のフレームの動体の残像 (動体が直前のフレームに存在した領域) が含まれる．以下に，本節の処理のステップを述べ，概要を図 3.3 に示す．

Step1. 入力映像において各フレーム，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 $\{v_j\}_{j=1\sim6}$ を確率変数 x とおいて，式 (3.1) の x が背景モデルに属する確率密度と所定のしきい値 θ の大小を式 (3.3) で比較する．

$$p(x) \quad > \quad \theta \tag{3.3}$$

Step3. ST-Patch 特徴量 $\{v_j\}_{j=1\sim6}$ の全要素で式 (3.3) 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 $\{v_j\}_{j=1\sim6}$ の中で一つでも式 (3.3) 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，閾値 π 個以上あれば動体有の候補フラグを ON にする．π の値は，事前に動体に想定する大きさに応じて定めておく．

Step5. 動体有の候補フラグの ON が，閾値 ρ 以上のフレームで連続すれば発報を出力する．閾値 ρ の値は，事前に **Step4** の判定の確からしさから定めておく．

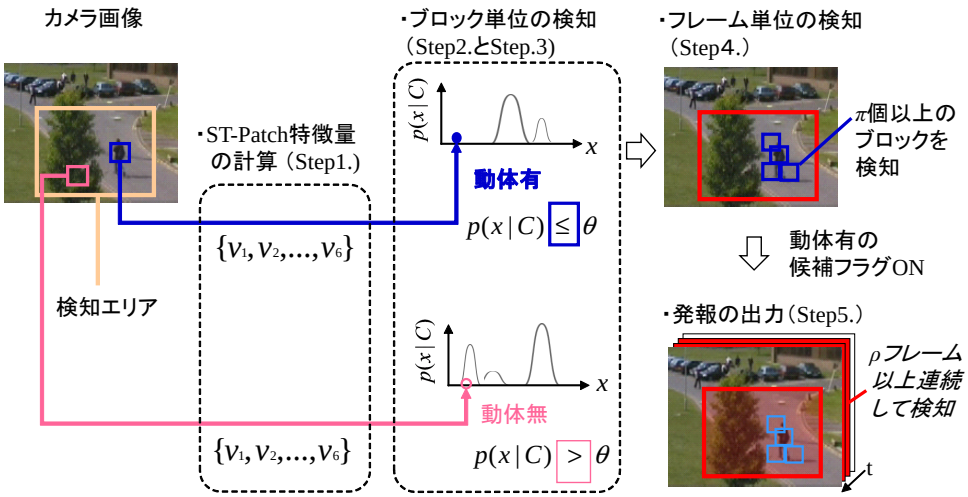


図 3.3: ST-Patch 特徴量による背景モデルを用いた動体検知の概要．

3.3.3　背景モデルの更新

本節では，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．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，背景の変化の仕方が時間毎に異なる．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隈なく収集しておくことは困難である．そこで，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常的な場合，時間経過と共に変化が消滅したり，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 2 つ以上の変化は冗長であり 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．図 3.4 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 $\{v_j\}_{j=1\sim6}$ が背景に属するか否かを，3.3.2 節の **Step2** の方法で判定する．

Step2a. **Step1** で背景に属すと判定したブロックでは，確率変数 x に最も近い i 番目の正規分布の重み係数 w_i，平均 μ_i，標準偏差 σ_i を，所定の更新率 α と式 (3.4) および式 (3.5) で更新する．

$$\begin{aligned} {w_i}' &= (1-\alpha)w_i + \alpha \\ {\mu_i}' &= (1-\eta)\mu_i + \eta x \\ \sigma_i^2 &= (1-\eta)\sigma_i^2 + \eta(x-\mu_i)^T(x-\mu_i) \end{aligned} \tag{3.4}$$

$$\eta = \alpha p_i(x) \tag{3.5}$$

Step2b. **Step1** で背景に属さないと判定したブロックでは，正規分布の数 M が最大値 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，σ_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_j の総和が 1 になるように正規化する．

以上の背景モデルの更新が有効に機能するためには，新たな背景の変化がある程度の時間幅で継続することが条件となる，これは新たな変化に対して，**Step3** で正規分布を追加した後に，その正規分布の重み係数を **Step2a** で何度か更新して大きくしなければ，式 (3.3) の背景モデルに属す条件の成立に寄与しないためである．また，以上の背景モデルの更新が有効に機能するには，動体に対して大きな重み係数の正規分布が出現しないように，動体の移動速度に対して十分小さく更新率 α を適切に設定する必要がある．

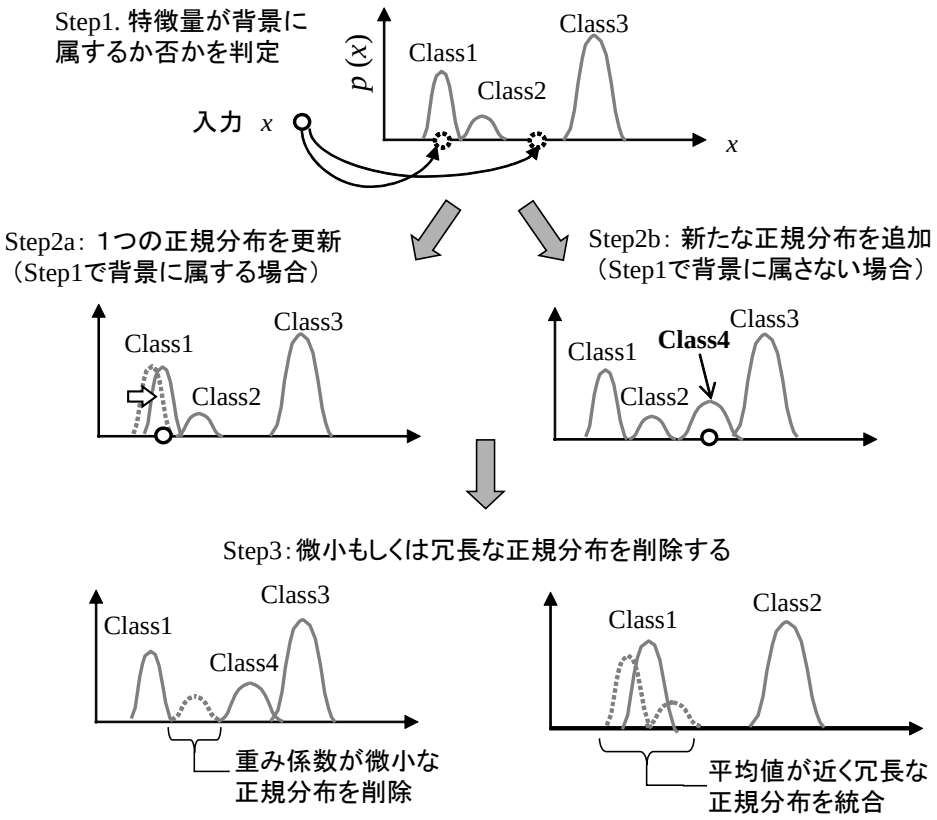


図 3.4: 背景モデルの更新の概要.

3.4 実験条件

本節では，評価実験の条件について述べる．評価実験では，提案手法に加えて，見えの特徴量を用いた従来手法として，正規化距離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，3.3.2 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 (3.3.2 節の **Step4** および **Step5**) は，ST-Patch 特徴量による手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 (3.4) と同じ更新率 α による Infinite Impulse Response(IIR) フィルタで更新した．なお，更新率 α は 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 × 縦 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，背景モデルの生成時に 3，更新時に最大 10 とした．

定量評価では，まず評価映像の各フレームにおいて，動体が十分な大きさで存在するフレームを

動体有，残りのフレームを動体無とし，動体の有無と発報の有無を比較して，True Positive(TP，動体有で発報有)，True Negative(TN，動体無で発報無)，False Positive(FP，動体無で発報有)，False Negative(FN，動体有で発報無) を判定した．次に，全フレームの判定の結果から，式 (3.6) で計算した Recall と Precision を評価指標とした．加えて，Recall と Precision の総合評価として，両者の調和平均で計算される F-Measure(F 値) を評価指標とした．

$$\begin{aligned}
\text{Recall} &= \frac{\sum \text{TP}}{\sum \text{FN} + \sum \text{TP}} \times 100 \\
\text{Precision} &= \frac{\sum \text{TP}}{\sum \text{FP} + \sum \text{TP}} \times 100 \\
\text{F-Measure} &= \left(\frac{\text{Recall}^{-1} + \text{Precision}^{-1}}{2} \right)^{-1}
\end{aligned} \tag{3.6}$$

評価映像は，屋外，屋内，エスカレータの 3 シーンを対象とした．以下，各シーンの評価映像について説明する．

■ 屋外のシーンの評価映像の概要

1 つ目の評価実験では，PETS2001[57] 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となっている．こ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，これが画像上の大域的な変化となっている．

この評価映像では，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．また樹木の近くを通る人を動体検知の対象とした．また，この評価映像には，学習用と評価用の 2 つの映像が用意されているので，学習用の映像を用いて背景モデルを生成し，評価用の映像を用いて動体検知の性能を評価した．背景モデルの生成では，学習用の映像のうち検知エリア内に人がいない 1,179 フレームを抜粋して生成に用いた．図 3.5 に，この背景モデルの生成に用いた映像のサンプルを示す．この評価映像では，評価用の映像の全 6,255 フレームのうち，人が十分な大きさで映っている 1,359 フレームを動体有，それ以外のフレームを動体無とした．

なお，樹木の揺れの強さは，図 3.5 に示す背景モデル生成用の映像では樹の葉だけが揺れる程度の弱い揺れしかほぼ出現しないが，動体検知の評価用の映像では樹の葉に加えて樹の枝が揺れる強い揺れが出現していた．

図 3.5: 屋外のシーンの背景生成に用いた映像のサンプル．

■ 屋内のシーンの評価映像の概要

2 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を対象にした [58]．この評価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．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，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．

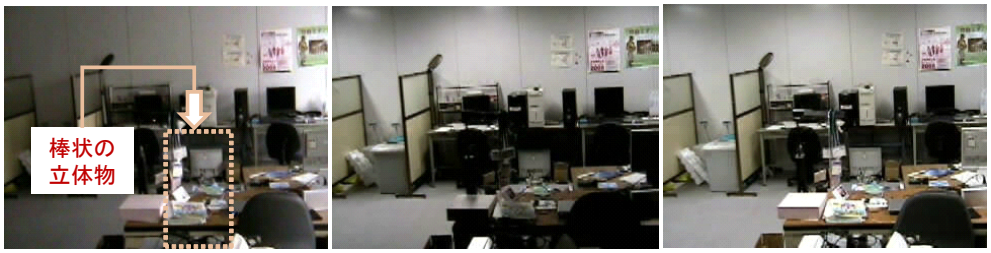


図 3.6: 屋内のシーンの背景生成に用いた映像のサンプル．

この評価映像では，映像の前半 665 フレームの人が進入しない部分から背景モデルを生成した．図 3.6 に背景モデルを生成した映像のサンプルを示す．図 3.6 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．そこで，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，検知エリアを横切る人を動体検知の対象とした．

この評価映像では，評価映像の後半 1,981 フレームを評価対象とし，検知エリア内に人が十分な大きさで映っている 99 フレームを動体有，それ以外のフレームを動体無とした．

■ エスカレータのシーンの評価映像の概要

3 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．この評価映像中のエスカレータには屋外に面した大きな窓が傍にあり，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，これが画像上の大域的な変化となっている．また，ステップに乗客がい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．ステップ上の乗客の影はステップの運行速度に応じて移動し，これが画像上の局所的な変化となっている．

この評価映像では，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 3.7 に，背景モデルを生成した映像のサンプルを示す．

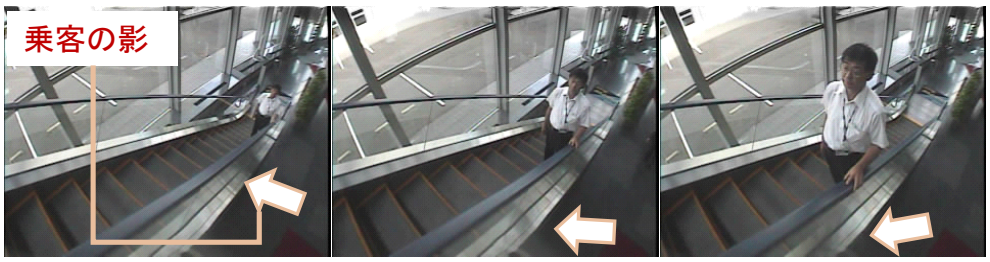


図 3.7: エスカレータのシーンの背景生成に用いた映像のサンプル．

この評価映像では，評価映像の全 3,250 フレームのうち，乗客の身体が十分な大きさではみ出た 215 フレームを動体有，それ以外のフレームを動体無とした．乗客が乗り出す挙動には，乗り出す部分を頭部，腕部，上体のようにバリエーションを持たせた．また，影のでき方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．

3.5　実験結果

本節では実験結果として，動体検知の例，定量評価の結果，提案手法中の構成要素の評価結果について述べる．

3.5.1　動体検知の例

■ 屋外のシーン

図 3.8 に屋外のシーンの評価実験の結果の代表例を示す．図 3.8 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 3.8 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の有無と発報の有無の比較結果を記した．

図 3.8: 屋外のシーンの結果の例.

図 3.8 の (1) は，風が強くなり樹木の葉の揺れが徐々に大きくなりだしたフレームである．図 3.8 の (2) は図 3.8 の (1) 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枝までが揺れ出したフレームである．図 3.8 の (3) は，図 3.8 の (2) の 289 フレーム後である．図 3.8 の (4) は，図 3.8 の (3) から後に急激に日照が弱くなったフレームである．なお，図 3.8 の (1) から (4) のいずれにおいても，動体 (人) は存在しない．

図 3.8 において，正規化距離による手法では，(1) から (4) で 4 つの False Positive が発生した．一方，ST-Patch 特徴量による手法では (1)(3)(4) で 3 つの True Negative と (2) で 1 つの False Positive という正規化距離より良好な結果を得た．ST-Patch 特徴量による手法では，(2) から (3) へフレームが進むにつれて，背景を誤検知したブロックの数は減少し，False Positive のフレームの発生頻度も減少した．この False Positive の減少は，3.3.3 節の背景更新の効果と考えられる．

図 3.8 の代表例に示す以外の傾向としては，ST-Patch 特徴量による手法は人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手法は，人が樹木に近いとき，および人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．

■ 屋内のシーンの実験結果の例

図 3.9 に屋内のシーンの評価実験の結果の代表例を示す．図 3.9 において，画面中央の四角形は検知エリアを示す．

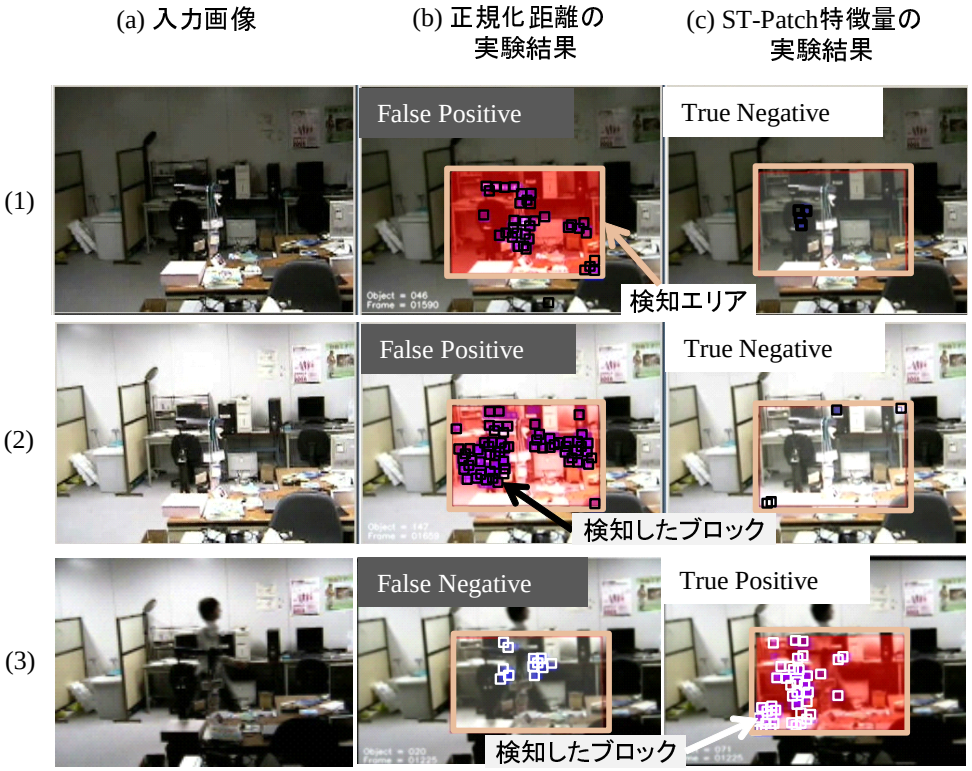


図 3.9: 屋内のシーンの結果の例．

図 3.9 の (1) は画面中央の付近を照らす天井照明以外の照明が全て消灯した直後である．図 3.9 の (2) は 図 3.9 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 3.9 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が横切った場面である．

図 3.9 において，正規化距離による手法では，(1) と (2) で 2 つの False Positive と (3) で 1 つの False Negative の結果となった．一方で，ST-Patch 特徴量による手法では，(1) と (2) で 2 つの True Negative と (3) で 1 つの True Positive という良好な結果を得た．なお図 3.9 の (3) において，ST-Patch 特徴量による手法では，3.3 節で述べた動体の残像を検知する特性によって，左から進入した人が棒状の構造物に遮蔽される以前の残像の部分 (図中の人の左傍) のブロックを多数検知することで，良好に True Positive の結果を得た．

図 3.9 の代表例以外の傾向として，ST-Patch 特徴量による手法は，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，正規化距離による手法は，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が発生していた．

■ エスカレータのシーンの実験結果の例

図 3.10 にエスカレータのシーンの評価実験の結果の代表例を示す．図 3.10 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を示す．

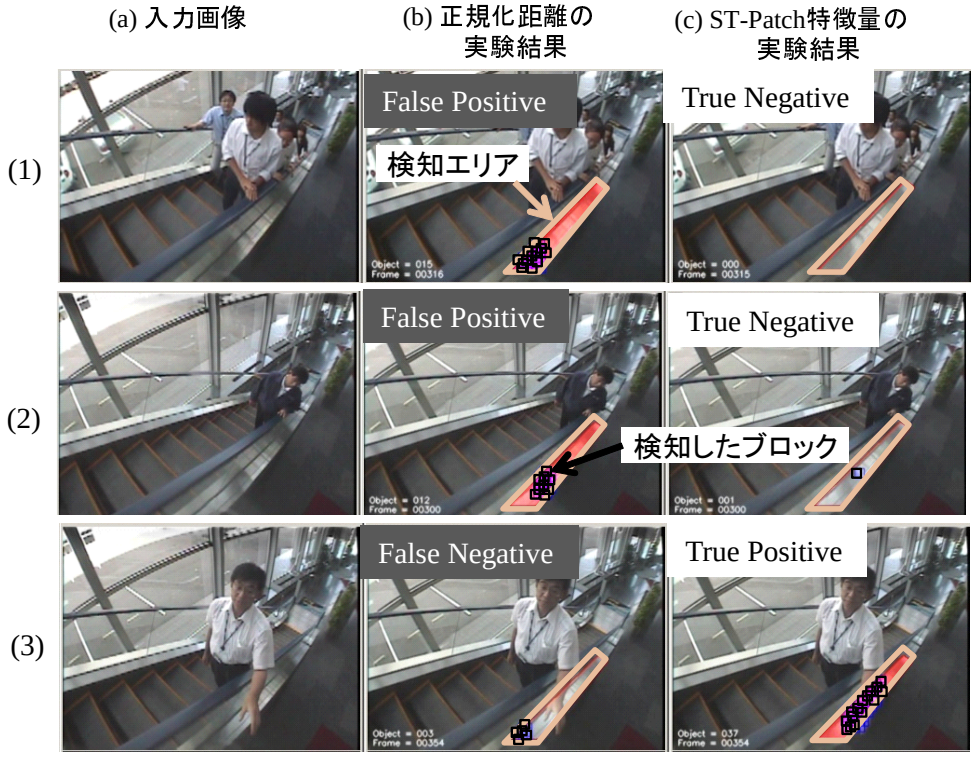


図 3.10: エスカレータのシーンの結果の例．

図 3.10 の (1) は複数の乗客が乗り出すことなく通過した場面である．図 3.10 の (2) は乗客が検知エリア付近に到達する以前の場面である．この (1) と (2) において，正規化距離による手法は背景生成時と動体検知時の間で乗客の数や乗客の姿勢が違うことによって，発生する影の形状が異なっていた．図 3.10 の (3) は，乗客の腕がはみ出したフレームであり，乗客の腕の明度は背景の金属部と近く，両者のコントラストは低くなっていた．

図 3.10 において，正規化距離による手法では，(1) と (2) で 2 つの False Positive と (3) で 1 つの False Negative の結果となった．一方で，ST-Patch 特徴量による手法では，(1) と (2) で 2 つの True Negative と (3) で 1 つの True Positive という良好な結果を得た．

図 3.10 において，ST-Patch 特徴量による手法が (1) の影の誤検知を抑止と (3) の乗り出した乗客の検知を両立しているのは，時空間中の濃度の勾配の強度や方向に両者間の差異を捉えたからである．ただしの代表例よりも乗客の数が多い場合や乗客の姿勢の変化が大きい場合，乗客の影が背景モデルの生成時のものから大きく乖離して，影の誤検知の抑止と乗り出した乗客の検知の両立が困難なケースがあった．

なお，エスカレータのシーンのように動体が一定速度で検知エリアを通過するシーンでは，動体の ST-Patch 特徴量が通過する間に均一に近い値をとる．そこで，3.3.3 節の背景モデルの更新では，動体の特徴量が背景モデルに高い重み係数で混入しないように，更新率 α を十分に小さく設定する必要がある．

3.5.2　定量評価の結果

■ 動体検知の例の定量評価結果

表 3.2 に 3.4 節の 3 シーンの実験結果の例から計算した 3 つの評価指標 (Precision, Recall,F 値) および定量評価時の π(3.3.2 節の **Step4** を参照) と ρ(3.3.2 節の **Step5** を参照) の値を示す．この π と ρ は，屋外のシーンにおいて検知結果が良好になるように設定した後，他のシーンでは必要に応じて調整を加えた．表 3.2 より 3 つのシーンのいずれにおいても，ST-Patch 特徴量による手法が正規化距離による手法よりも，評価指標全てで上回るとともに，特に Precision と F 値で大幅に優れていたことがわかる．表 3.2 の上から順に Precision で 38%，68%，36%，F 値で 32%，55%，32%，前者が後者より優れる．3 つのシーンにおいて画像上の背景には様々な種類の局所的ならびに大域的な変化が出現することを考えると，ST-Patch 特徴量による手法は正規化距離による手法よりも頑健に背景の変化を吸収可能と考えられる．

表 3.2: 動体検知の例の定量評価結果.

	正規化距離 [%]			提案手法 [%]			パラメータ	
	Recall	Precision	F 値	Recall	Precision	F 値	π	ρ
屋外	76	41	53	92	79	85	10	3
屋内	64	26	37	91	94	92	20	1
エスカレータ	78	34	47	90	70	79	10	3

■ Precision Recall 曲線による評価

3.4 節の 3 つの評価映像を対象に，定量評価として Precision Recall 曲線を求めた．Precision Recall 曲線は，ある手法の複数の動作点における Precision と Recall を結んだグラフである．Precision Recall 曲線では，グラフが右上にあるほど，その手法が優れた認識性能を発揮したことを示している．

評価対象の手法は，3.5.2 節で述べた (i)ST-Patch および (ii) 正規化距離を用いた手法と，(iii) 小領域の単位の更新型 Radial Reach Correlation (RRC) と画素単位の Parzen 推定の 2 つの背景モデルの複合 [52] を用いた手法の合計 3 つ手法とした．(iii) は，大域的および局所的な背景の変化に頑健な従来手法として評価した．

(iii) の背景モデルは，画素単位で動体を検知するため，検知エリア内で検知した画素の総和が閾値を超えることを，フレーム毎の検知の条件とした (3.3.2 節，**Step4** 参照)．(iii) の背景モデルのパラメータは，RRC の背景の更新率 α を (i) と (ii) と共通で 0.01 とした以外，文献 [52] 中の値を用いた．

Precision Recall 曲線上の複数の動作点を求めるには，(i) の手法では ST-Patch 特徴量の閾値 (式 (3.3) の θ，(ii) の手法では正規化距離の閾値を変化させた．また，(iii) の手法では，更新型 RRC および Parzen 推定それぞれの画素単位の検知の閾値を組み合わせで変化させた．動作点が十分な範囲に分布しない場合は，さらに補助的に (i) と (ii) の手法では検知したブロックの数の閾値 (3.3.2 節の **Step4** の π) を，(iii) の手法では検知した画素の数の閾値を変化させた．

なお，(i) と (ii) の手法では複数の動作点を求めるために閾値は 10 通り変えた．一方，(iii) の手法では，2 つの閾値を 5 通りずつ変えた組み合わせで 25 通り変えた．(i) と (ii) の手法は，(iii) の手法よりも 10 ／ 25 の少ない回数で特性を測ることができた．

Precision Recall 曲線の計算結果を図 3.11 に示す．まず (i) および (iii) の手法と (ii) の手法を比較すると，3 つのシーンいずれにおいても，前者のグラフは後者のグラフよりも常に右上にあることがわかる．特に屋外と屋内のシーンで顕著である．これは，局所的な背景の変化に対して頑健な特性を持つ前者と持たない後者の差と考えられる．

次に，(i) と (iii) の手法を比較すると，3 つのシーンいずれでも，(i) の手法のグラフが (iii) の手法のグラフよりほぼ常に右上にあるが，屋外との屋内のシーンでの差は小さいが，エスカレータのシーンでの両者の差は大きい．定量的に F 値の最大値で比較すると，屋外と屋内とエスカレータのシーンの順番に，(i) の手法が 88%，93%，87%，(iii) の手法が 83%，89%，73%，両者の差が 5%，4%，14%であった．

ここで，エスカレータのシーンでは，(i) と (iii) のどちらの手法も，ステップの移動により一定速度で検知エリアを通過する乗客の影の誤検知が強く影響している．(iii) の手法は Recall が 100%付近では (i) の手法と同等の Precision を発揮するが，誤検知を低減して Precision を向上するには，多数の失報の増加が伴い Recall が大きく低下する．その一方で (i) の手法では，Precision の向上に伴う Recall の低下は小さい．これは，(i) の手法の時空間の特徴量よる背景モデルが，エスカレータの乗客の影のような一定速度で動く背景の誤検知の抑止に有効性を持つためと考えられる．

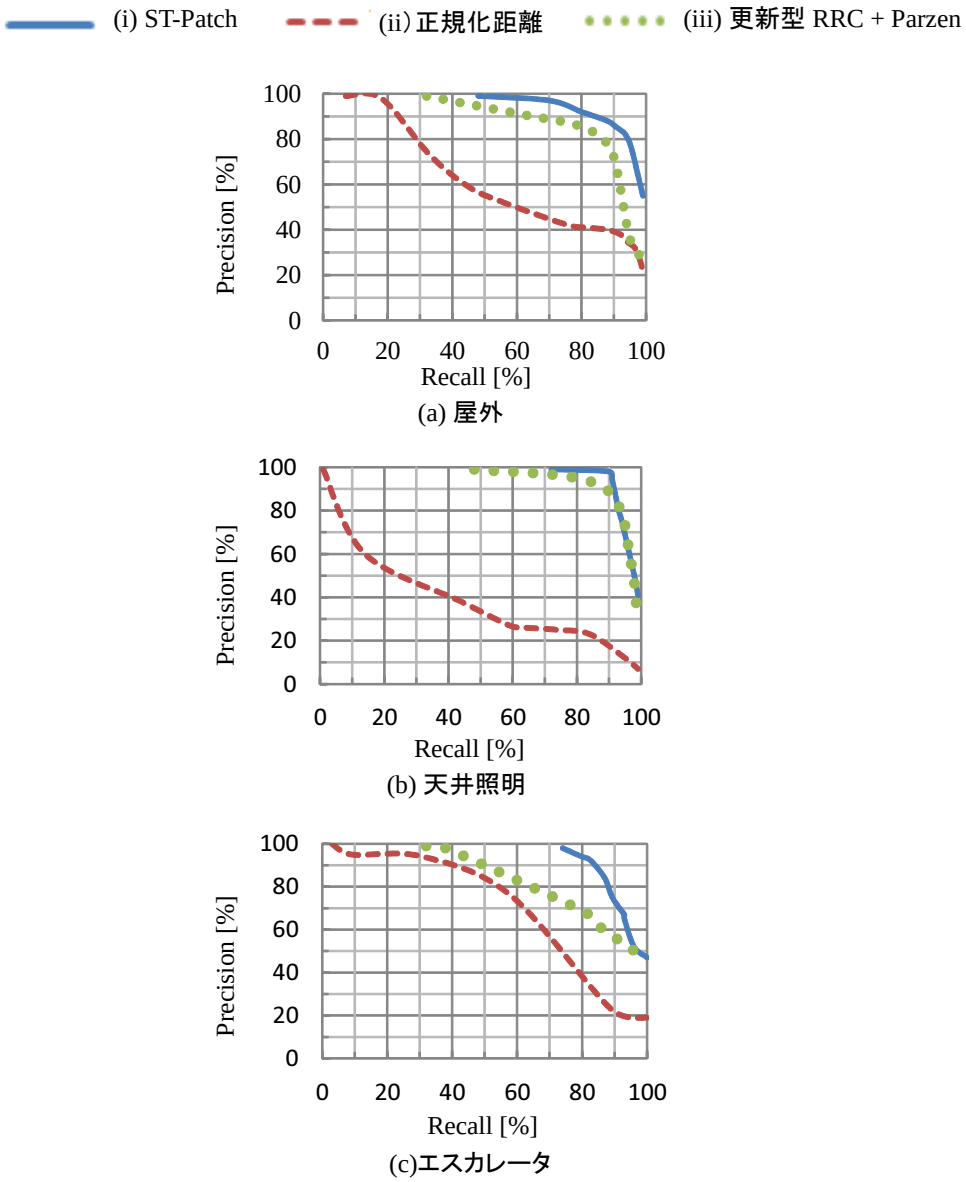


図 3.11: Precision Recall 曲線による評価.

3.5.3 提案手法の構成要素の評価結果

本節では，提案手法の 3 つの構成要素の有効性の有無の評価，および有効性を発揮する条件を評価した結果について述べる．ここで提案手法における 3 つの構成要素とは，(1)ST-Patch 特徴量による動きの情報の抽出 (3.3.1 節)，(2) 混合正規分布による背景モデルの生成 (3.3.2 節)，(3) 動体検知中の背景モデルの更新 (3.3.3 節) である．3.5.2 節の評価実験では，最終的な動体検知の結果を評価

するため，3 つの構成要素の内どの要素が有効であるのか，またどの要素がどのような条件で有効であるのかが不明である．そこで本節の評価実験は，各構成要素の評価を以下の手順で行う．

- 提案手法から 1 つの構成要素の特質を排除した 3 つの比較手法を作成する．
- 提案手法と 3 つの比較手法の間で，Precision Recall 曲線を比較する．比較の結果から，各構成要素の有効性の有無を評価する．
- Recall を同一に揃えた条件で，背景の変化の状態別の False Positive の発生数を提案手法と比較手法とで比較する．比較の結果から，各構成要素が有効性を発揮する条件を評価する．

■ 提案手法の構成要素の評価実験

提案手法の 3 つの構成要素それぞれに対応した比較手法を以下の様に設定した．

- **比較手法 1**(特徴量の動きの情報を排除)：
  背景モデルの生成と動体の検知において，ST-Patch 特徴量から動きの成分 (式 (2.4)) の第 4〜6 項) を排除して，見えの成分 (式 (2.4) の第 1〜3 項) のみを使用する．
- **比較手法 2**(背景モデルの多峰性を排除)：
  混合正規分布の代わりに単峰のガウス分布で背景モデルを生成する．また，背景の更新において，新たな正規分布を追加しない (3.3.3 節の **Step2b** 参照)．
- **比較手法 3**(背景モデルの更新を排除)：
  背景モデルを動体の検知の間に更新しない．

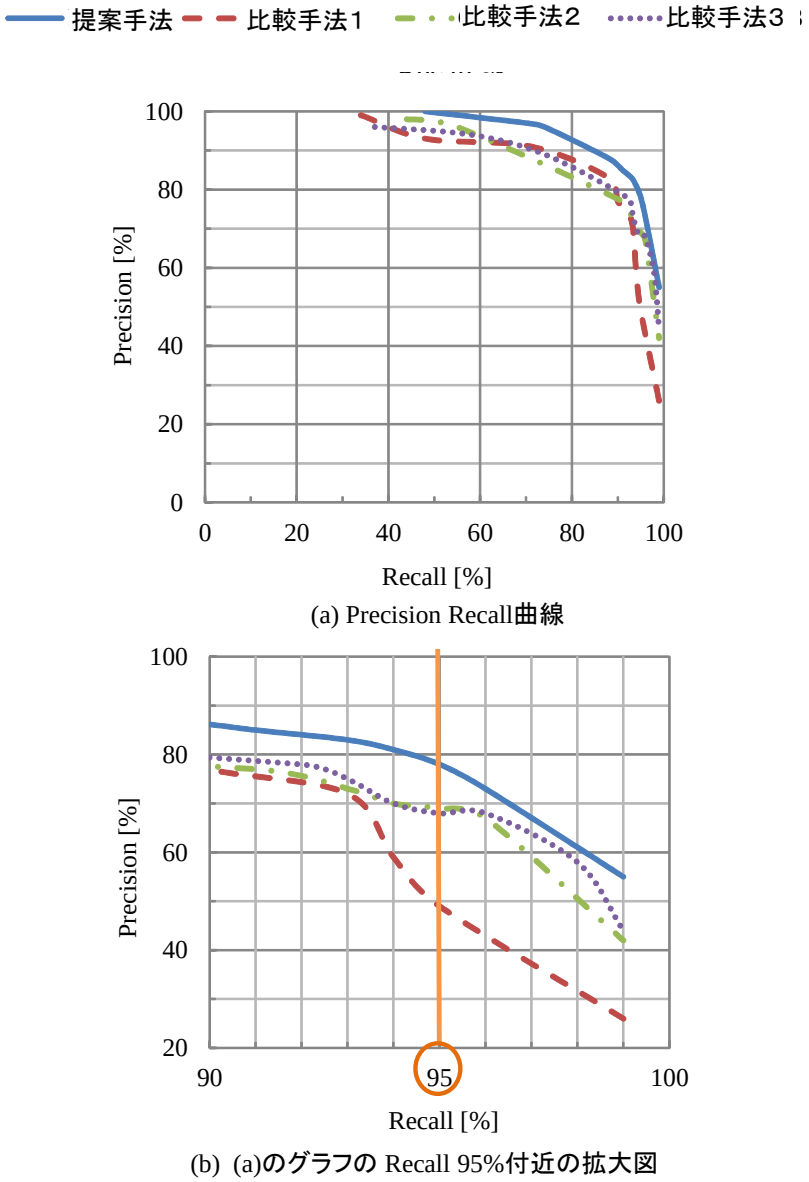
■ 構成要素の有効性の評価

屋外の評価映像に対する Precision Recall 曲線を図 3.12(a) に示す．図 3.12(a) より，Precision が 100%付近から Recall が 100% 付近までの広い範囲に渡って，提案手法が 3 つの比較手法よりも常に右上側にあり優れた認識性能を発揮したことがわかる．よって，提案手法の 3 つの構成要素は，いずれも背景の変化の吸収に有効であることがわかる．特に (b) の拡大図のように Recall が 90% ∼ 100%の高い検知率の領域に着目すると，Recall が 90 ∼ 93%の間はどの比較手法も Precision が同程度だが，Recall が 93 ∼ 100%の区間では比較手法 1 の Precision が他の比較手法よりも低い．これは後者の区間において，比較手法 1 に対応した構成要素 1 の効果が大きいことを示す．

■ 構成要素が有効性を発揮する条件の評価

屋外の評価映像を対象に，Recall の条件を揃えた動作点において，背景の変化の状態別の False Positive を提案手法と 3 つの比較手法とで比較した．Recall の動作点には，95%を選択した．この

95%は，Recall を過度に高く設定すると動体検知の感度が過敏になり偶発的なノイズによる False Positive が両者の差に影響する傾向があったことを考慮してバランスを取り設定した動作点である．



(a) Precision Recall曲線

(b) (a)のグラフの Recall 95%付近の拡大図

図 3.12: 提案手法と比較手法の Precision Recall 曲線.

一方，背景の変化の状態は，屋外の評価映像における大域的な変化である日照の強度の変化の有／無と，同映像の局所的な変化である樹木の揺れの有／無でラベル付けした．ただし，樹木の揺れは強弱で変化の様子が大きく変わるので，変化があるときには，弱い揺れ (樹の葉だけが揺れる程度) と，強い揺れ (樹の葉に加えて樹の枝や幹が揺れる程度) に細分した．

以上の手順で求めた提案手法ならびに比較手法 1〜3 の背景の変化の状態別の False Positive の発生数を表 3.3 に示す．表 3.3 において，比較手法 1〜3 の左列は False Positive の絶対数，右列は提案手法と比較した差分である．表 3.3 において，日照の変化の ANY の列は，日照の変化が有る時と無い時の合計である．

表 3.3 において，ある条件を変える前後に着目した時に，False Positive の差分が顕著に増加する部分を下線と太字で示す．このとき．3 つの比較手法いずれにも．顕著に増加する部分が見られた．これは提案手法の 3 つの構成要素が，特定の条件にて特に有効性を発揮することを示す．以下にその条件について述べる．

表 3.3: 提案手法と比較手法の誤報要因．

樹の揺れ	日照の変化	提案手法	比較手法 1		比較手法 2		比較手法 3		フレーム数
無	無	0	0	差無	0	差無	0	差無	1,693
	有	0	0	差無	0	差無	0	差無	872
弱	無	13	309	+296	37	+24	23	+10	1,046
	有	22	321	+299	46	+24	50	+28	666
強	無	138	389	+251	213	+75	234	+96	1,319
	有	124	282	+158	241	**+117**	229	+105	800
無	ANY	0	0	差無	0	差無	0	差無	2,565
弱	ANY	35	630	**+595**	83	+48	73	+38	1,712
強	ANY	262	671	**+409**	454	**+192**	463	**+201**	2,119
総数		297	1,301	1,004	537	+ 240	536	+239	6,396

- **構成要素 1**(特徴量の動きの情報)

  比較手法 1 の実験結果において，樹の揺れが無い時には日照の変化の有無によらず False Positive の差は無い．一方，揺れの強弱によらず樹の揺れがあるときは False Positive が大幅に増加する (日照の強度の変化の有無の合計で，樹の揺れが弱で+595，強で+409)．以上より，特徴量の動きの成分は，樹の揺れのような局所的な変化の吸収に大きく有効である．

- **構成要素 2**(背景モデルの多峰性)

  比較手法 2 の実験結果において，樹の揺れが強いときに日照の強度の変化が無い時よりも (+75) 有るとき (+117) の方が False Positive の増加が大きい．以上より，背景モデルの多峰性は，多様な背景の変化の吸収に有効である．

  また，日照の強度の変化の有無の合計で，樹の揺れが弱い時 (+48) よりも強い時 (+192) の方が False Positive の増加が大きい．以上の比較結果に強い樹の揺れが背景モデルの生成時に出現しないことを加えると (3.4 節を参照)，背景モデルの多峰性は新たな背景の変化への追従に有効である．

- **構成要素 3**(背景モデルの更新)

  比較手法 3 の実験結果において，日照の強度の変化が有る時と無い時の合計は，樹の揺れが弱い時 (+38) よりも強い時 (+201) の方が False Positive の増加が大きい．以上の実験結果に強い樹の揺れが背景モデルの生成時に出現しないことを加えると (3.4 節を参照)，背景モデルの更新は新たな背景の変化への追従に有効である．

3.6 まとめ

本章では，見えと動きの情報を併せ持った時空間特徴量の 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．また評価実験により，局所的および大域的な変化が背景に存在する屋外，屋内，エスカレータの 3 つのシーンにおいて，提案手法が動きの情報を持たない見えの情報のみを用いた従来手法よりも優れる実験結果を示した．具体には，屋外，屋内，エスカレータの 3 つのシーンにおいて，提案手法の F 値は，正規化距離より順に 32%，55%，32%優れ，更新型 RRC と Parzen 推定を用いた手法よりも順に 5%，4%，14%優れた．

また，提案手法と，3 つの構成要素 (特徴量の動きの情報，背景モデルの多峰性，背景モデルの更新) を排除した比較手法との実験結果の比較より，上記の 3 つの構成要素が全て背景の変化の吸収に有効であること，および各構成要素が有効性を発揮する条件を明らかにした．構成要素は順に，樹の揺れのような局所的な変化，多様な背景の変化ならびに新たな背景の変化への追従，新たな背景の変化への追従にそれぞれ特に有効である．

第4章

時空間特徴量の統計的学習を用いた異常動作検出

3章では，カメラ画像中の背景の大域的および局所的な変化に対して頑健な動体検知の手法について述べた．しかしながら，この手法では検知エリア単位で動体を検知するために，検知エリア内の物体の種類 (人) や物体の動作といった詳細な情報を得ることができない．本章では，時空間特徴量の統計的学習により，カメラ画像中の人領域を検出し，人領域内の動きの非定常度から転倒という突発的な動作を検出する手法について述べる．

提案手法は，時空間特徴量を用いた非定常度の計算，時空間特徴の統計的学習による人領域の検出，人の部分に重み付けした非定常度による異常動作検出という3つの要素から構成される．時空間特徴量にはST-Patch特徴量を用い，統計的学習にはReal Adaboostを用いる．

評価実験では，背景中でステップが定常的に動き続けるエスカレータのシーンにおいて，提案手法は従来手法よりも優れた検出率を取得し，提案手法の有効性を確認した．また，屈み動作の実験結果と転倒の実験結果を比較することで，提案手法が動きの小さな動作の検出に優れていることを確認した．

4.1 従来手法の課題

異常動作検出の先行研究として，文献 [59] では 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が提案されている．また，文献 [60] では，動画像の特徴の共分散行列を用いた手法が提案されている．先行研究の中で最も代表的な手法として，文献 [61] では，Cubic Higher-order Local Auto-Correlation(CHLAC) を用いた手法が提案されている．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位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．文献 [61] では，CHLAC と部分空間法の組み合わせで異常動作を検出する手法が提案されている．先ず学習のフェーズでは，正常動作のサンプルから CHLAC の部分空間を構築する．次に検出のフェーズでは，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動作として検出する．CHLAC は異常動作の他，各種のイベントシーンからのイベントの検出への適用が提案されている [14].

他にも，フローベクトルを使った方法も先行研究の中で代表的なアプローチである．文献 [19] では，エレベータ内においてフローベクトルを算出し，その個数ならびに方向と強度の分散から暴れの異常動作を検出する手法が提案されている．文献 [27] では，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動作として検出する手法が提案されている．

異常動作検出をより多種類のシーンに適用するためには，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 やフローベクトル等の従来手法は，この動的な背景への対応は困難である．これは，背景の動きの特徴量の中に人の動きの特徴量が埋もれてしまうからである．この問題は，異常動作の画像上の動きが小さいほど顕著となる．

4.2 提案手法のアプローチ

本章では，異常動作の中で特に突発的なものを認識対象とし，動的な背景下において突発的な異常動作を検出する手法を提案する．突発的な異常動作とは，歩行中の転倒や階段の昇降中の踏み外しのような正常動作から逸脱しておきる異常動作である．提案手法では，まず物体の見えと動きの情報を併せ持った時空間特徴の ST-Patch 特徴量を用いて，背景および人の見えと動きの特徴抽出および動きの非定常度を計算する．次に，ST-Patch 特徴量の統計的学習により矩形単位で人領域を検出する．これにより，矩形単位で，背景と人を大まかに区分する．最後に，人領域内の非定常度に人領域検出の信頼度による重み付けを行う．これにより，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．図 4.1 に提案手法の流れを示す．

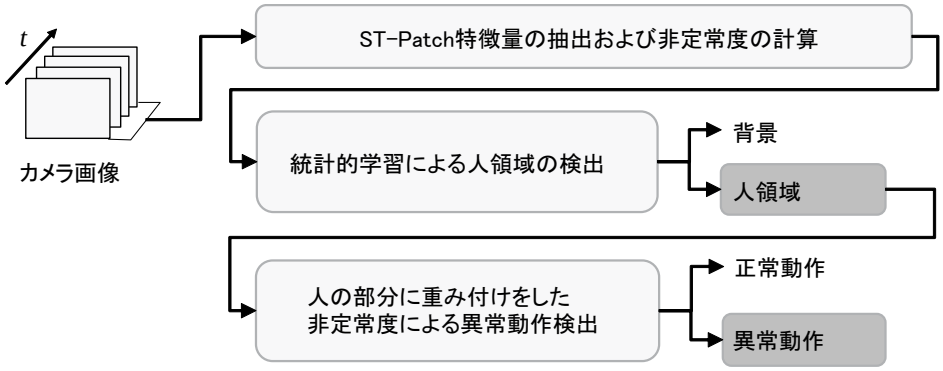


図 4.1: 提案手法の流れ.

本章では，以下，4.3 節にて ST-Patch 特徴量と人領域検出の統計的学習を用いた異常動作検出の手法について述べる．4.4 節にて，エスカレータを動的な背景，転倒を異常動作とした評価実験の実験条件について述べる．4.5 節にて，評価実験の実験結果について述べ，提案手法の有効性を示す．4.6 節にて，まとめを述べる．

4.3 ST-patch 特徴量と人領域検出の統計的学習を用いた異常動作検出

本節では，ST-patch 特徴量を用いた非定常度の計算，統計的学習による人領域検出，人の部分に重み付けをした非定常度による異常動作検出について述べる．

4.3.1 ST-patch 特徴量を用いた非定常度の計算

ST-Patch 特徴量 [7] は，2.2.2 節で述べたように，動画像の小領域中の時空間の濃度の勾配値から見えと動きを捉えた特徴量である．提案手法では，式 (2.4) に示した ST-Patch 特徴量の 6 次元の要素を用いる．また，提案手法では，ST-Patch 特徴量から計算される Continuous Rank-Increase Measure(CRIM)[7] と呼ばれる動きの非定常度を，転倒等の突発的な異常動作の発生の指標として用いる．以下，この CRIM について説明する．

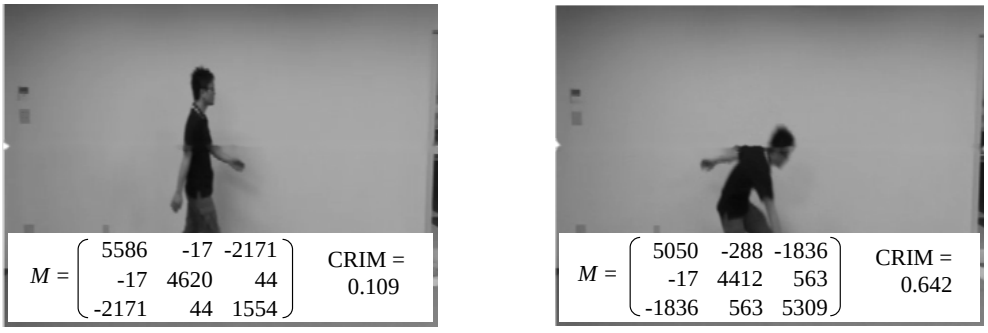
式 (2.3) の行列 M の固有値を $\lambda_1 \geq \lambda_2 \geq \lambda_3$ とし，行列 M の上 2×2 を行列 $M^\diamond$ とし，行列 $M^\diamond$ の固有値を $\lambda_1^\diamond \geq \lambda_2^\diamond$ とすると，$\lambda_1 \geq \lambda_1^\diamond \geq \lambda_2 \geq \lambda_2^\diamond \geq \lambda_3$ の関係にある．これらの固有値から，CRIM は式 (4.1) の Δr で計算する．

$$\Delta r = \lambda_2 \lambda_3 / \lambda_1^\diamond \lambda_2^\diamond \qquad (0 \leq \Delta r \leq 1) \tag{4.1}$$

照明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．物体が静止もしくは一律な

並行移動を継続する場合は，Δr は 0 に近付く．反対に，物体の動きが変化する場合は，Δr は 1 に近付く．

図 4.2 の例を用いて，CRIM に正常動作の継続と異常動作の発生を区分できる特質があることを示す．この例では動画像全体から CRIM を計算した．図 4.2 (a) は画面の左端から右端まで通過した正常動作である．図 4.2(b) は左端から右端に通過する途中で転倒した異常動作である．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．



(a) 通過（正常動作）　　(b) 転倒（異常動作）

図 4.2: CRIM の数値例.

4.3.2　統計的学習による人領域検出

提案手法では背景の動きの影響を軽減するために，大まかに矩形単位で人領域を検出し，人領域内の時空間の勾配値から CRIM を計算する．以下，この人領域検出について述べる．

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-like 特徴量の統計的学習による顔領域の検出 [62] や，HoG 特徴量の統計的学習による人領域の検出 [4] 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[63] という手法を用いる．この手法は学習時に，特徴量の要素毎に 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 4.3 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 クラスを正常動作時の人の領域，Negative クラスを背景およ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，人の進入も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取ってしまう．高い値の CRIM は，4.3.3 節で述べる異常動作の検出に悪影響するので抑止をした．

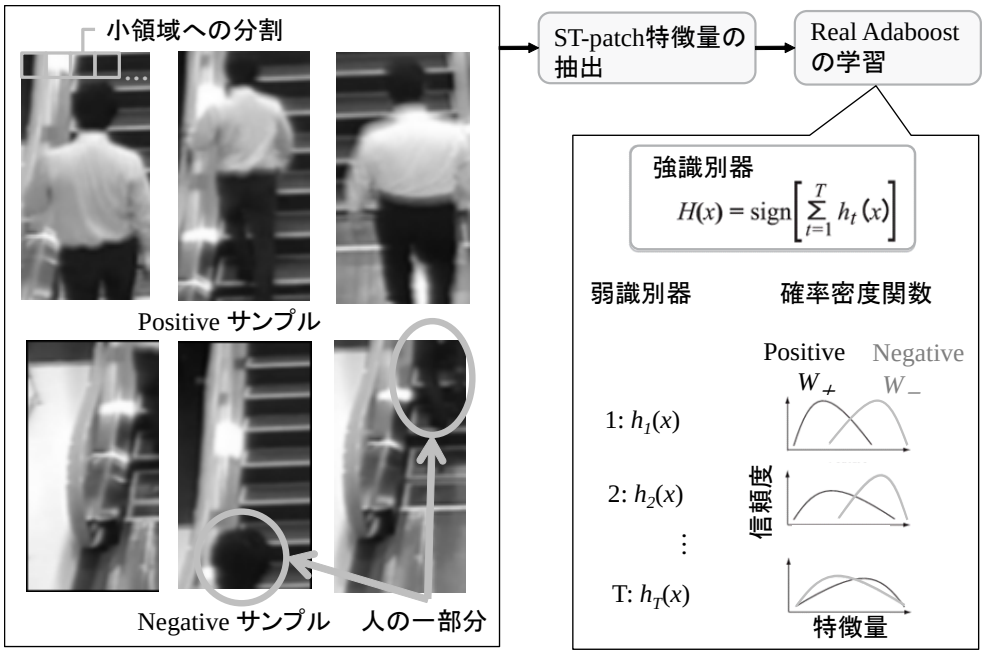


図 4.3: 人領域検出の統計的学習の概要.

人領域の検出は，学習結果を用いて次の流れにより行う．まず，入力映像の各フレームを小領域の集合に分割して，小領域毎の 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閾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である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] ×5[frame] および間隔は 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 4,446(=741×6) である．

4.3.3 人の部分に重み付けをした非定常度による異常動作検出

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勾配値から CRIM を計算する．この重み付けには，人領域内の小領域毎の 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．この重み付けには，以下に述べる統計的学習の信頼度を用いる．

統計的学習の信頼度は，先ず弱識別器毎に求め，次に小領域毎に求める．弱識別器の信頼度は Real Adaboost の応答値と分離度の 2 つの指標で計算する．指標の 1 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 および 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 および 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 (4.2) の $h(x)$ となる．

$$h(x) = \frac{1}{2} \ln \frac{{W_+}^j + \varepsilon}{{W_-}^j + \varepsilon} \tag{4.2}$$

式 (4.2) 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取り，反対に小さければ負の符号を取る．また，$h(x)$ が正の時，W_+^j の値が W_-^j の値より大きいほど，$h(x)$ は大きな値を取る．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 4.3 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．

2 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 (4.3) により計算する．

$$\alpha = 1 - \sum_i \sqrt{{W_+}^i {W_-}^i} \tag{4.3}$$

式 (4.3) の α は 0 から 1 の値を取る．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取る．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 4.3 の場合，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．

以上求めた応答値 $h(x)$ と分離度 α は，式 (4.4) 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．

$$e(x) = \begin{cases} \alpha\, h(x) & \text{if } h(x) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.4}$$

次に，小領域単位の信頼度 E は，小領域内の弱識別器の信頼度 $e(x)$ の総和である式 (4.5) で計算する．この小領域の分割は，4.3.2 節で述べた人領域の検出と同一である．

$$E = \sum_{x \in S} e(x) \tag{4.5}$$

式 (4.5) 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程に大きくなる．

信頼度で荷重した CRIM は，以下の様に計算する．まず信頼度を乗じた濃度勾配 G_E を式 (4.6) で計算する．

$$G_E = \left[\begin{array}{cccc} E_1 {G_1}^T & E_2 {G_2}^T & .... & E_B {G_B}^T \end{array} \right]^T_{n \times 3} \tag{4.6}$$

式 (4.6) 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勾配値のスタックである ($k = 1, 2 \ldots B$)．G_E の例を図 4.4 に示す．G_E を求めた後は，4.3.1 節の手順で 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ことが可能となる．

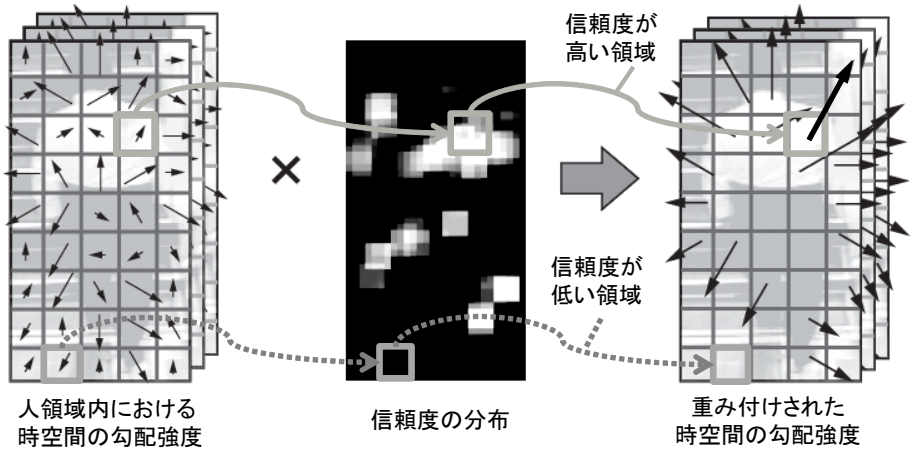


図 4.4: 信頼度で重み付けした時空間の勾配値の例.

提案手法では，各フレームにおいて全ての人領域から CRIM を計算する．そして，1 つ以上の人領域内において CRIM が閾値を超えることを条件に，フレーム単位で異常動作の発生を検出する．

4.4 実験条件

本節では評価実験の条件について述べる．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．本実験における異常動作は，エスカレータの乗降時の転倒を対象とする．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ことが報告されている [64]．本実験における転倒のデータは，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正面からやや見下ろす角度で撮影した．このカメラで撮影する画像中において，人の形状は真後から見たように映り，これは人領域検出によって良好な条件である．人の寸法は，奥行によって手前と奥側とで大きさが変化する．カメラは 360×240[pixel]×30[fps] の画像を時間的に連続して取得し，これを評価実験の入力とする．

評価の対象は，提案手法 2 つと，従来手法 2 つとした．提案手法は以下の 2 つである．

- 提案手法 1：信頼度の重み付き CRIM による手法
- 提案手法 2：信頼度の重み無し 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 4.3.3 節の方法で，提案手法 2 は 4.3.1 節の方法で CRIM を計算する．CRIM を計算した後の異常動作の検出は，どちらも 4.3.3 節の末尾で述べた方法で行う．各フレームで，人領域内の CRIM の最大値が 1 つでも閾値を超過すればすれば異常動作有，それ以外の場合ときは異常動作無と判定する．

従来手法は，背景の動きの影響の抑止と異常動作の検出に有効な従来の技術を適用した以下の 2 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 [65] 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，以下の通りである．提案手法 1 と 2 の人領域の検出の学習には，Positive クラスを 2,002 個，Negative クラスを 6,048 個のサンプルを用いた．従来手法 1 と 2 は，CHLAC の累積時間は，予備実験から文献 [61] 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 3,000 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．

4.5 実験結果

本節では，実験結果の例，定量評価の結果について述べる．また，提案手法の詳細な特性の評価結果として，人の大きさの影響の評価結果，人領域検出の評価結果，時空間特徴量の比較評価の結果について述べる．

4.5.1 実験結果の例

■ 提案手法の実験結果の例

提案手法の実験結果の例として，CRIM のグラフと代表フレームを図 4.5 に示す．グラフより，提案手法 1 は異常動作のフレームで閾値より大幅に高く，提案手法 2 は異常動作のフレームで閾値を僅かに超える場合と超えない場合があることがわかる．

図 4.5 の 100，308，686，982，1,267 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 2 つは正常動作，後 3 つは異常動作である．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，画像上の動きが大きい．686 フレームの転倒と比較すると，982 および 1,267 フレームの転倒の画像上の動きは小さい．それぞれ，位置が奥側およ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 は正常動作と異常動作を全て正しく判定した．一方，提案手法 2 は，982 と 1,267 フレームの転倒が未検出であった．

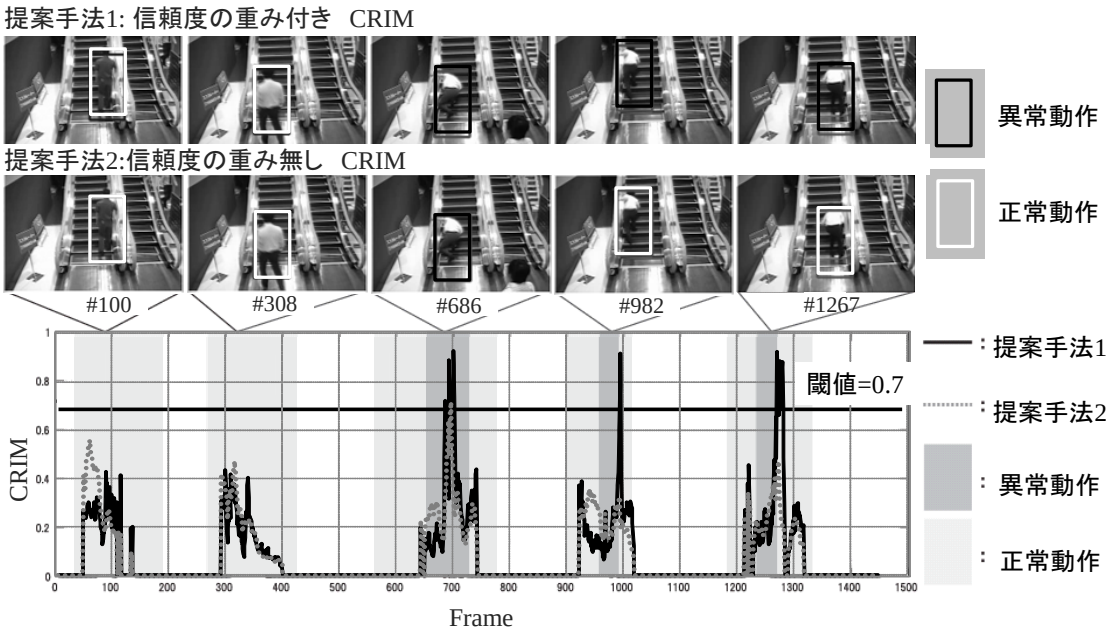


図 4.5: 提案手法の実験結果の例．

ここで図 4.5 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 4.5 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，人の部分内の画素数と人領域全体内の全画素数の比は，図 4.5 の左から順に 39, 59, 41, 31, 35%，5 フレームで平均で 41%であった．これらの比 41%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，平均では信頼度を使わない時より 26%優れた．

■ 従来手法の実験結果の例

まず，図 4.6 に CHLAC の入力のフレーム間差分の 2 値画像の例を示す．640 フレームはステップに近付く途中，1,250 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，人の輪郭を大凡で捉えている．一方，背景を含んだ右側は，人の輪郭に加えて，ステップの動きを捉えている．他のフレームも同様の傾向にあった．

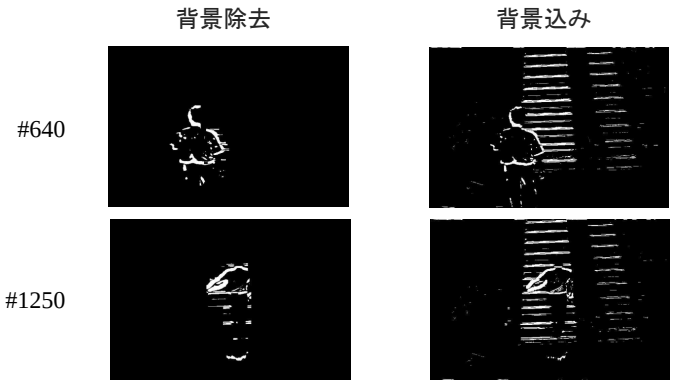


図 4.6: 2 値化したフレーム間差分の例.

次に，図 4.7 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動作は図 4.5 と同一である．逸脱度のグラフより，従来手法 1 は正常動作と異常動作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無いことがわかる．代表フレームにおいて，従来手法 1 は 282 フレームの誤検出と 982 フレームの未検出があった．前者は，横切りの発生がカメラに近く，画像上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，画像上の動きが小さいためと考える．従来手法 2 は 282 フレームが誤検出の上，転倒は 3 つどれも未検出である．

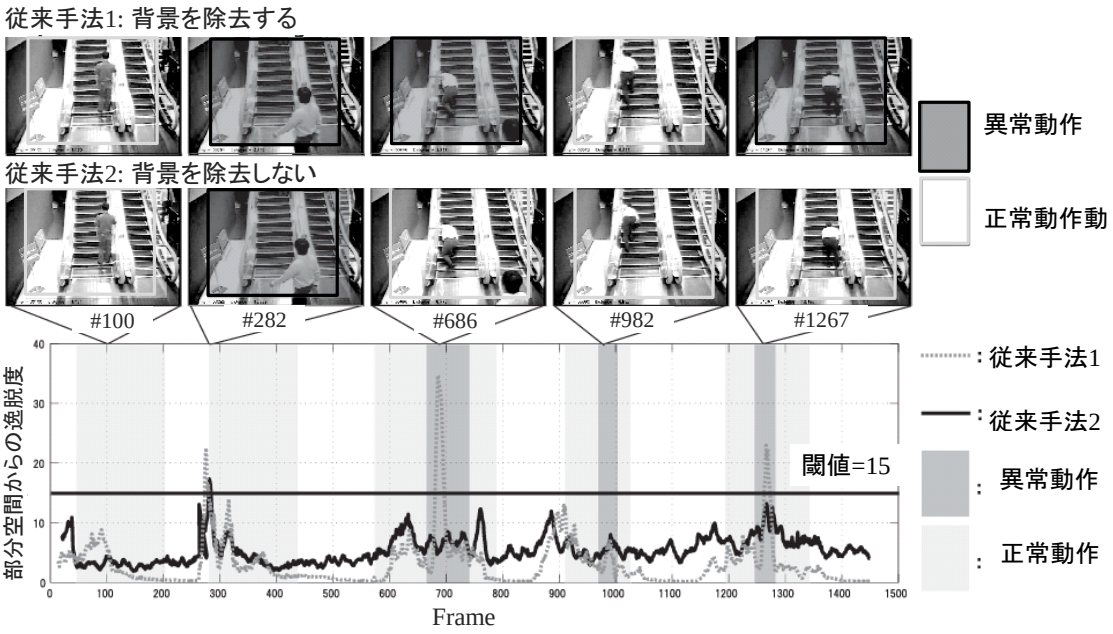


図 4.7: 従来手法の実験結果の例.

4.5.2　定量評価の結果

定量評価は転倒と屈みの 2 つの動作を対象とした．屈みは，図 4.8 のようにステップの前まで歩いて近付いた後に，直立から一息に屈む動作である．転倒より見た目の動きが大きい．動作の種類が与える影響を調査するために評価の対象とした．

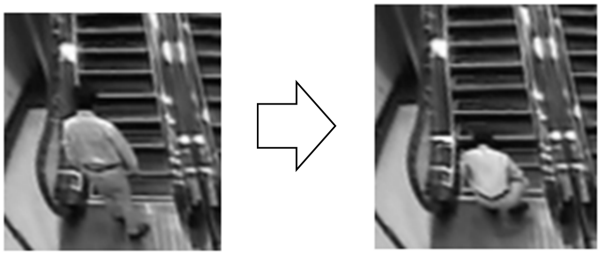


図 4.8: 屈みの画像の例．

転倒の評価には，正常動作 18 ケース，異常動作 12 ケースを用いた．屈みの評価には，正常動作 18 ケース，異常動作 9 ケースを用いた．正常動作は，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．屈みは，屈む位置とステップの方向を変化させた．以上の転倒と屈みおよび正常動作の評価映像のフレーム数は合計 6,486 フレームである．各手法で異常動作検出の閾値を変えたときの検出率と誤検出率を，図 4.9 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程に性能が優れる．図 4.9(a) の屈みは手法間の差が小さい．一方，図 4.9(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．

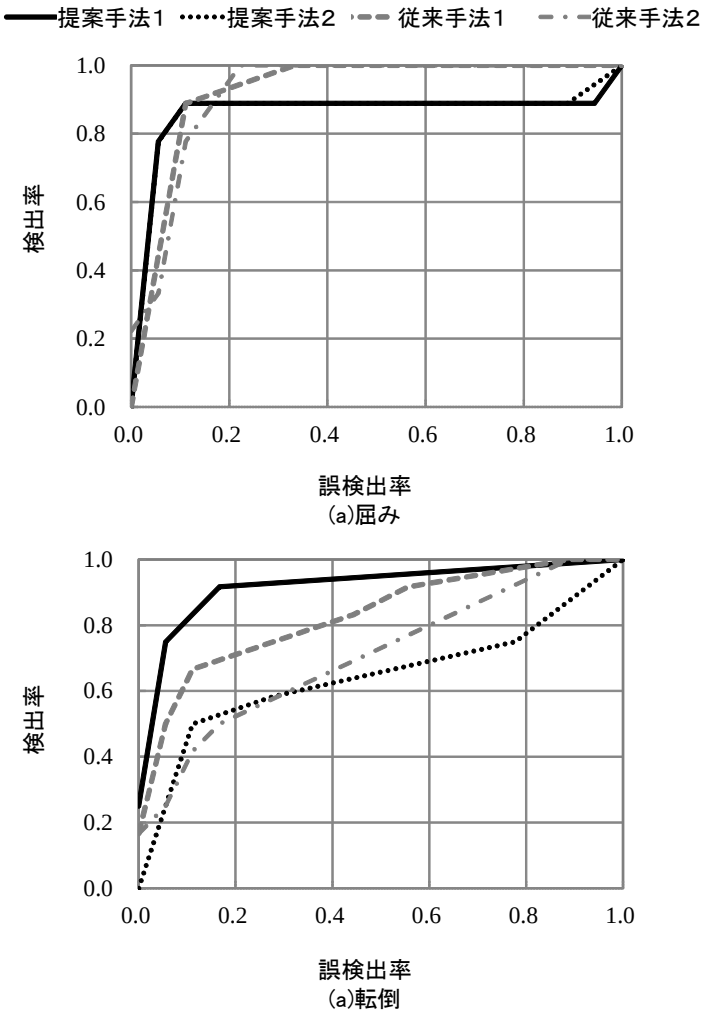


図 4.9: 転倒動作と屈み動作の ROC カーブ.

表 4.1 に示す誤検出率 10%を代表点に取った検出率を比較する．先ず屈みは，差が小さい．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，見えが大きな異常動作は，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く，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 の屈みと転倒の評価結果と，転倒の方が屈みよりも画像上の動きが小さいことを合わせると，前者は後者より画像上の動きが小さな異常動作の検出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 4.5.5 節の評価結果を含めて述べる．

定量評価を通じて，検出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 4.5.1 節の実験結果の例と同様であった．画像上の動きが小さな異常動作の未検出，および画像上の動きが大きな

表 4.1: 誤検出率 10%時の異常動作の検出率.

手法	提案手法 1	提案手法 2	従来手法 1	従来手法 2
特徴量	CRIM		CHLAC	
背景変動の抑止法	人領域検出と信頼度重み付け	人領域検出	背景差分	無
屈み	89%	89%	89 %	78%
転倒	83%	59%	67%	42%

正常動作の誤検出が主だった．提案手法 1 には，通過の途中の不意の動作への誤検出があった．ステップの手前の振り返りが一例である．これは，異常動作を検出した後に時空間特徴量等から別途の動作認識をして，非定常度の高い正常動作と異常動作を区分すれば解決が可能と考える．また，検出した人領域の手前を，他の人が横切る瞬間の動きに誤検出があった．これは，複数の人領域間の遮蔽を判定し，手前の人の動きの影響を抑止すれば解決が可能と考える．なお，複数人が通過する途中に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出できた．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蔽されなかったためと考える．人相互の遮蔽が頻繁な状況，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等の身体の一部を捉える方法の適用で可能と考える．

4.5.3 人の大きさの影響の評価結果

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動作の検出率で評価した．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%時の奥と手前の検出率の内訳を表 4.2 のように求めた．表 4.2 において，まず提案手法 2 の手前と奥の検出率を比べると，両者で差が無い．これは信頼度の重みの無い非定常度は，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 の検出率を比べると，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．

表 4.2: 提案手法の転倒の検出率の内訳.

	提案手法 1	提案手法 2
手前	50%	25 %
奥側	33%	25 %
合計	83%	50 %

4.5.4 人領域検出の評価結果

提案手法中の人領域の検出率と誤検出率を評価した．対象データは，4.5.2 節の実験の延べ 6,486 フレームとした．検出率で要求条件は，十分に高くて異常動作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 4.5 のグラフのように非定常度が異常動作中に変化することと，非定常度が閾値を超えた部分を逃すと異常動作が検出できないことに依る．この要求条件が満たされないときは，異常動作の検出率は低下する．特に，検出率がゼロの時は，非定常度が一切抽出されないため，異常動作を全く検出できない．

表 4.3 に検出率の実験結果を示す．異常動作の種別 (転倒／屈み)，人の大きさ (手前／奥，4.5.3 節と同じ基準)，動作の種別 (正常動作／異常動作) の条件が影響することを想定して，条件毎に平均値 (全体の傾向) と最低値 (最も不良な場合) を代表値で求めた．異常動作時は，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，正常動作時の人領域検出は異常動作の検出に関与しないが，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 4.3: 人領域の検出率.

		手前		奥側	
		正常動作	異常動作	正常動作	異常動作
転倒	平均	96%	89%	89%	87%
	最少	64%	56%	53%	48%
屈み	平均	88%	80%	-	-
	最少	58%	**4%**	-	-

表 4.3 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%，屈みで 4%であった．転倒の 48%は，誤検出率 10 %時に提案手法 1 が該当ケースで異常動作を検出できたので，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，屈みの 4%は，提案手法 1 と 2 共に誤検出率が 100%近くでないと該当ケースで異常動作を検出できなかったので，今回の実験では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屈んだことと，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，屈みの最低値を除けば，異常動作時の検出率はどの項目でも，正常動作時から 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，転倒ならびに屈む時の人とを，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らない．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，異常動作時でも大きくは性能が劣化しないと考えられる．

次に，誤検出率での要求条件は，十分に低く異常動作の偶発的な誤検出を頻発しないことである．偶発的な誤検出とは，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%，屈み 11%であった．共に頻繁な異常動作の誤検出の誘発は無く，前述の要求条件を満たしていた．

4.5.5 時空間特徴量の比較評価の結果

提案手法中の CRIM と従来手法中の CHLAC を時空間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 および 2 は，背景の動きの影響の抑止の方法が提案手法 1 および 2 と同じである．比較手法 1 および 2 の CHLAC では，累積時間は従来手法と同じ 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人が正常に通過する 3,000 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 および 15 とした．実験結果として，転倒の異常動作の誤検出率 10%時の検出率を表 4.4，ROC カーブを図 4.10 に示す．

表 4.4: 誤検出率 10%時の異常動作の検出率の比較．

手法	提案手法 1	提案手法 2	比較手法 1	比較手法 2
特徴量	CRIM		CHLAC	
背景変動の抑止法	人領域検出と信頼度重み付け	人領域検出	人領域検出と信頼度重み付け	人領域検出
転倒	83%	50%	75%	50%

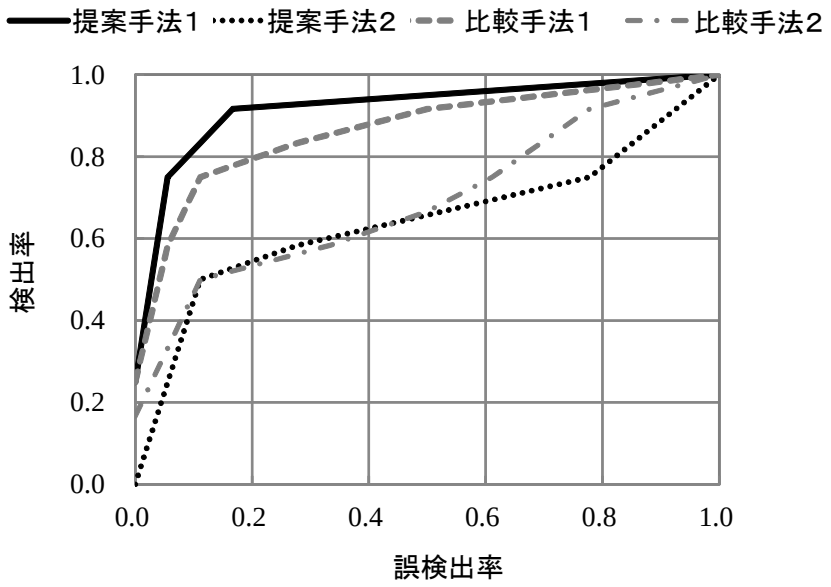


図 4.10: 提案手法と比較手法の ROC カーブ.

提案手法 2 と比較手法 2 の検出率が，表 4.4 で共に 50%であることと，図 4.10 のグラフで誤検出率 10%〜50%の間でほぼ一致することより，信頼度の重み付けが無いと CRIM と CHLAC の差はほぼ無いと考える．一方で，提案手法 1 と比較手法 1 の検出率を比べると，前者は後者より表 4.4 で 8%優れ，図 4.10 のグラフで誤検出率 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動作の検出で優れることが影響していると考えられる．これは時空間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動作の検出に適すことに依る．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出が可能となる画像上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依る．また，転倒の内，異常動作の代表例である図 4.5 の 1,267 フレームの動きの小さな転倒 (転倒直後に手を突いて大きく姿勢の崩れを軽減が崩れることを抑止) の例では，提案手法 1 は誤検出率 10%時において提案手法 1 は検出できたが，比較手法 1 は検出できなかったことに依る (比較手法 1 は誤検出率 10%時では検出できずに，誤検出率 27%時でに検出できた).

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時は，誤検出率 10%時に図 4.5 の 1,267 フレームの転倒を検出できた．ただし，CHLAC に増えたノイズの影響を抑止するために閾値を低くする必要があったので，全体の検出率が 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動作の検出で優れると考えられる．

最後に，表 4.1 と表 4.4 に示す誤検出率 10%時の提案手法，比較手法，従来手法の検出率の平均値を表 4.5 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，時空間特徴量の差である．比較手法が従来手法より 9%優れるのは，時空間特徴量が共通な

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，時空間特徴量と背景の動きの影響の抑止の方法の差の両方に依り，内訳は前者が 4%，後者が 9%である．

表 4.5: 異常動作の検出率の平均値の比較．

手法	提案手法	比較手法	従来手法
特徴量	CRIM	CHLAC	CHLAC
背景変動の抑止法	人領域検出と (+ 信頼度重み付け)	人領域検出と (+ 信頼度重み付け)	背景差分 (or 無)
転倒	67%	63%	54%

4.6 まとめ

本章では，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動作を検出する手法を提案した．背景中でエスカレータのステップが動き続けるシーンにおける転倒の検出を対象とした評価実験では，提案手法は CHLAC とフレーム間差分を用いた従来手法と比べて，誤検出率 10%時の検出率が 16%優れる実験結果を取得した．この実験結果から，提案手法の有効性を確認した．

また，この評価実験のシーンにおける屈みと転倒の 2 種類の動作の検出の評価結果を比較することで，提案手法が特に画像上の動きが小さな異常動作の検出に有効であることを確認した．これは，CHLAC とフレーム間差分を用いた従来手法と提案手法の間において，画像上の動きが大きな屈みの動作では両者で差が無かったが，画像上の動きが屈みよりも小さな転倒では提案手法が従来手法よりも優れたことに依る．

第5章

距離画像の三面図展開を用いた生成型学習による動作認識

4章では，カメラ画像中の時空間特徴量の統計的学習により人領域を検出し，人領域内の動きの非定常度から突発的な動作を検出する手法について述べた．しかしながら，この手法では，人領域検出が人の全身の特徴量のパターンを統計的学習に学習するために，動作を通じて人の姿勢の変化が大きな時には対応できないという課題点がある．また，認識対象とする動作の種類が転倒のような突発的な動作に限られるという制約条件がある．本章では，距離画像センサから得られる距離画像から，幅広い種類の動作を対象とした人の見えの変化に対して頑健な動作認識手法を提案する．

提案手法では，学習時と認識時とでカメラ視点が変化するときに，人の向きの変化が動作認識に及ぼす影響を抑止する．まず，距離画像を人の向きの変化に伴う位置の変化に対して不変となる様にシルエットの三面図展開を行う．次に，学習時のカメラ視点の距離画像のサンプルから，カメラ視点を変化させた時の三面図を生成することで，生成範囲内のカメラ視点毎に応じた弱識別器を学習する．最後に，これら弱識別器と認識時と同じカメラ視点の少数のサンプルから最適な強識別器を構築する．これにより，学習サンプルの収集の手間を小さく抑えながら，カメラ視点が変化した時の人の動作を良好に認識する．

評価実験では，学習時と認識時とで画像中の人の向きが約 90° 変化する条件において，暴れや歩行といった基本的な動作を対象とした動作認識の精度が改善し，提案手法の有効性を確認した．

5.1　従来手法の課題

本節では，まずカメラ視点の変化が動作認識に及ぼす影響について述べる．次に，距離画像を用いた動作認識において，カメラ視点の変化に対応した従来手法およびその課題点について述べる．

5.1.1　カメラ視点の変化による影響

人を異なるカメラ視点から撮影すると，カメラから見た人の向きが変わることで，人の向きが画像中において変化する．この人の向きによる変化は，動画像中の見えや動きを捉える時空間特徴量を大きく変化させて，動作認識の精度を低下させる原因となる．また，人とカメラの距離が変化する場合には画像中における人の大きさも変化するため，マルチスケールのスキャニング等による対応策が求められる．

なお，同一個所に設置したカメラの視野内で，人の位置や向きが変わった時にも，カメラと人の相対的な位置関係が変わることより，画像中で人の向きは変化する．また，人の位置の変化に応じて人の遠近が変わるときには，画像中における人の大きさの変化も伴う．よって，本論文では，同一カメラの視野内における人の位置や向きの変化も，カメラ視点の変化に含める．

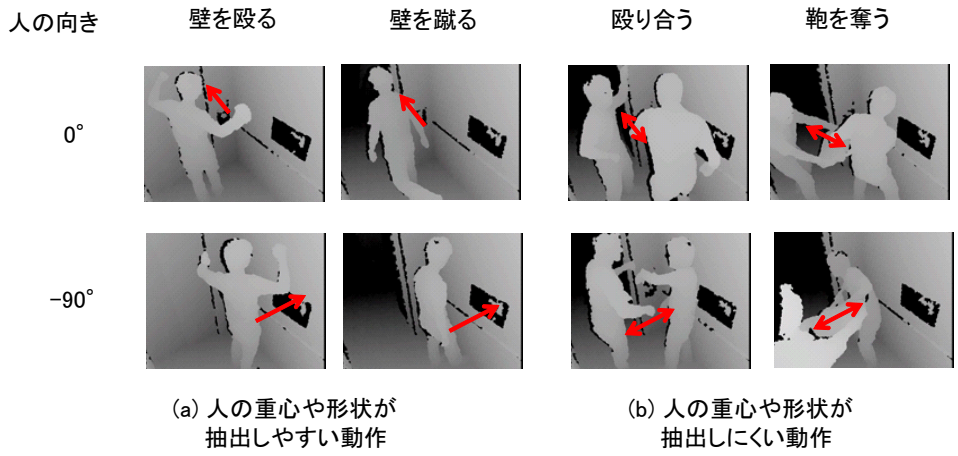


図 5.1: カメラ視点が変化した距離画像の例．

カメラ視点の違う距離画像の例を図 5.1 に示す．明るい階調の画素ほど距離画像センサに近く，暗い階調の画素ほど距離画像センサから遠いことを示す．第 1 行と第 2 行では 0° と −90° という 2 種類の向きにおける 4 種類の動作している距離画像を示す．ここで，0° の距離画像から各動作の時空間特徴量を求めて統計的学習を行い，−90° の距離画像から求めた時空間特徴量とこの統計的学習の結果を用いて動作認識をした場合について考える．この時，時空間特徴量あるいは統計的学習のいずれかで人の向きの変化を吸収しなければ，学習時と認識時との間の人の見えの変化が時空間特徴

量の変化を引き起こし，動作認識の性能は低下する．また，学習時の 0° よりも認識時の −90° の距離画像において，人が僅かに手前側に来て，画像中で大ききが異なるので，その差を吸収することが望ましい．

5.1.2 カメラ視点の変化に対応した従来手法とその課題

これまでに，距離画像の距離情報を利用することで，カメラ視点の変化に対応した時空間特徴量を動作認識に適用する手法が提案されている [41][33][66]．これらの従来手法を大きく分けると，人の向きを推定して人の向きに応じて時空間特徴量を補正する手法と，時空間特徴量を人の向きに対して不変化する方法の 2 つとなる．なお，いずれの従来手法も三次元座標で時空間特徴量を記述するため，画像中の人の大きさの変化も併せて不変化される．

前者のアプローチとして，文献 [41] では，距離画像中の人のスケルトンの認識結果を利用して，人の重心と向きを推定し，鉛直および人の正面を座標軸とした極座標 $(\phi - \theta)$ のヒストグラム中における関節位置の分布を時空間特徴量として抽出する手法が提案されている．この方法では，この極座標系が人の向きの変化に追従するために，カメラ視点の変化に対応することができる．

後者のアプローチとして，文献 [33] では，前処理として距離画像から人の重心の位置を検出し，距離画像のフレーム間差分の各画素の三次元座標を人の重心を原点とした極座標 $(\phi - \theta)$ のヒストグラムに集約した後に，ヒストグラムを球面調和関数で変換することで人の向きに不変化した時空間特徴量を抽出する手法が提案されている．また，文献 [66] では，前処理として同様に距離画像から人の重心の位置を検出した後に， Motion History Volume と呼ばれるボリュームデータ中で変化が起きた時刻を濃淡値で記録したデータを，前記の人の重心を原点とした円筒座標系の Fast Fourier Transform(FFT) のスペクトルに変換することで人の向きに不変化した時空間特徴量を抽出する手法が提案されている．

以上の従来手法では，距離画像から人の向きや重心といった人の形状に関わる特徴量が正確に抽出できることが前提となっている．ただし，この前提は複数の人が接触し合う動作や大きく姿勢を変える動作の認識には適さない．文献 [41] では人の向きと重心，文献 [33] と文献 [66] では人の重心を前段で検出する必要があり，図 5.1(b) に示す殴り合う動作や鞄を奪う動作では，複数の人が接触したり大きく姿勢が変化するため適用が困難である．また，これまでに述べた動作は，セキュリティシステムに適用する動作認識として強く求められる．

カメラ視点の変化に対応する方法としては，以上述べた従来手法以外にも，認識時のカメラ視点毎の学習サンプルを事前に収集しておき，この学習サンプルから抽出した時空間特徴量と教師信号（動作認識の正解値）から，カメラ視点毎に動作認識の学習を行う手法が考えられる．ただし，カメラ視点毎の学習サンプルを収集する方法では，高い精度で動作認識を行うために多くの学習サンプルが必要となるため，認識時のカメラ視点毎の学習サンプルの収集に手間を要するという課題がある．

5.2　提案手法のアプローチ

図 5.2 に提案手法のアプローチを示す．図 5.2 では，前処理として (1) において，学習時のカメラ視点における距離画像の学習サンプルを射影変換のモデルを用いて三次元データ (x, y, z) に変換した後，無限遠の仮想視点から見た三面図に展開する．この三面図には，カメラ視点が変化した時において，人の向きの変化に伴う人の位置の変化に対して不変という特長がある．次に (2) では，カメラ視点を所定範囲で網羅的に変化させた三次元データおよび三面図を生成する．ここでは，三次元データを利用することで仮想的に自由に視点を変化することが可能になるという特性を利用している．(3) では，視点毎に生成したデータから時空間特徴量を抽出し，網羅的に変化させたカメラ視点毎に動作認識の弱識別器を学習する．最後に (4) では，認識時のカメラ視点の少数の距離画像の学習サンプルから，認識時のカメラ視点に最適な強識別器を構築する．ここで，(1)(4) における学習サンプルとは，距離画像と動作認識の教師信号の組である．

提案手法では，以上のアプローチにより，認識時に動作する人の向きや重心等の人の形状に関わる特徴量を抽出する必要がないため，従来手法の課題である複数の人が接触したり大きく姿勢を変える状況に対応することが可能となる．また，認識時のカメラ視点の学習サンプルの数を小さく抑えることで，学習サンプルの収集の手間を小さく抑えることができる．

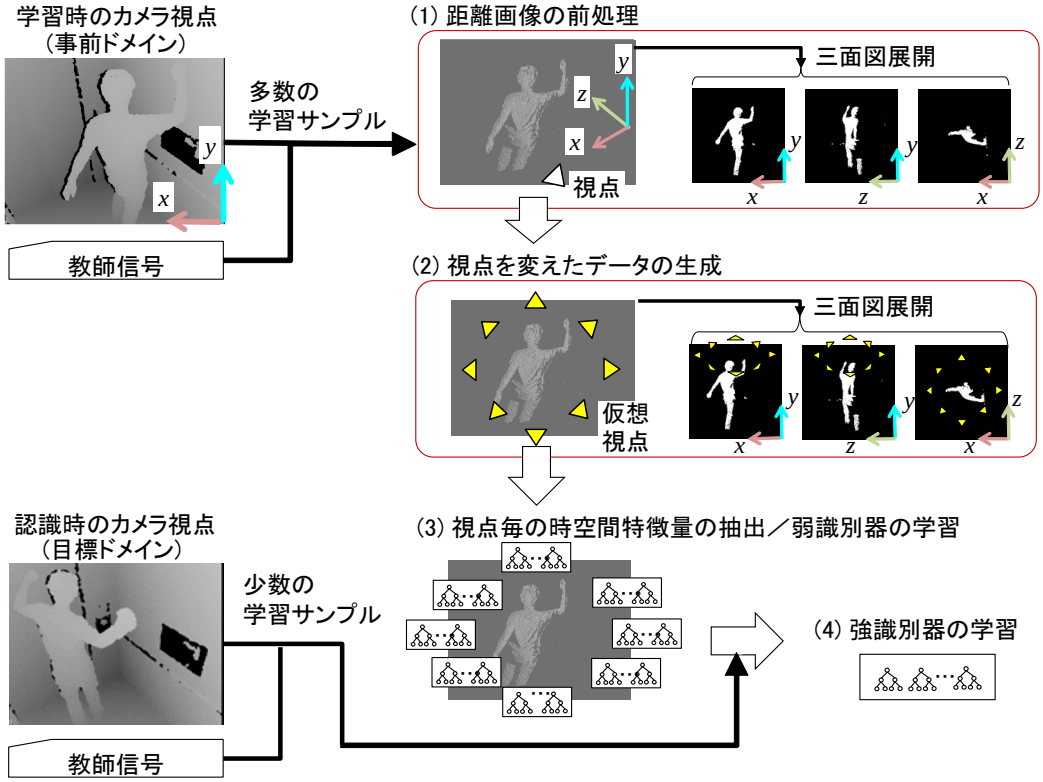


図 5.2: 提案手法のアプローチ．

なお，以上述べた提案手法は，学習時のカメラ視点を事前ドメイン，認識時のカメラ視点を目標ドメインとした転移学習の一種といえる．事前ドメインにおける動作認識の学習結果を利用することで，目標ドメインの動作認識の学習に必要な学習サンプルの量を小さく抑えることを図る．転移学習という言葉には様々な解釈があるが，機械学習の分野においては NIPS 2005 workshop - inductive transfer: 10 years later の論文募集要項で,「新規タスクの効果的な仮説を効率的に見つけ出すために，一つ以上の別のタスクで学習された知識を得て，それを適用する問題」と定義されている．この定義に従う転移学習は，事前ドメインとよばれる事前に学習した識別器や学習サンプルを，目的ドメインと呼ばれる新たな環境に対する識別器の構築に利用することで効率化を目指している．転移学習では応用例として，人検出における俯角の変化への対応 [67] や環境条件の変化への対応 [68] および新規サンプルの追加 [69] 等が提案されている．

以下，5.3 節では，距離画像の三面図展開を用いた生成型学習について述べる．5.4 節では，生成型学習の学習結果を用いた動作認識について述べる．5.5 節では，提案手法の評価実験の実験条件について述べる．5.6 節では，提案手法の評価実験の実験結果について述べる．5.7 節では，本章のまとめを述べる．

5.3 距離画像の三面図展開を用いた生成型学習

本節では提案手法の学習のフローを示した後，フロー中の各ステップの処理について述べる．学習のフローを図 5.3 に示す．

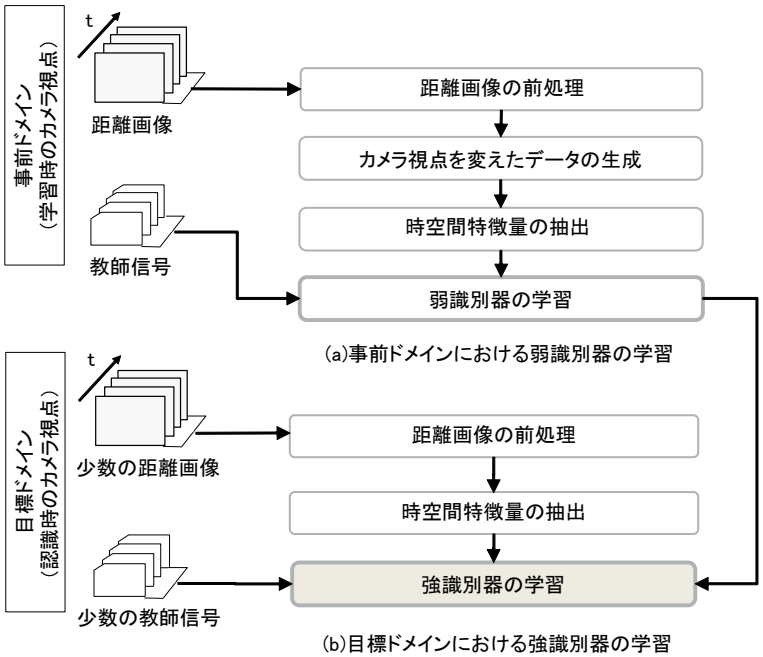


図 5.3: 提案手法の学習のフロー．

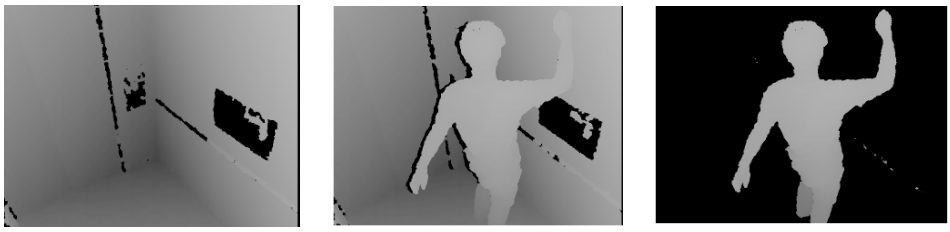
学習のフローは，図 5.3(a) の事前ドメインにおける弱識別器の学習と，図 5.3(b) の目標ドメインにおける強識別器の学習に分けられる．事前ドメインにおける弱識別器の学習 (図 5.3(a)) では，事前ドメインの距離画像および動作の教師信号を入力として，距離画像の前処理，カメラ視点を変えたデータの生成，時空間特徴量の抽出，弱識別器の学習を順に行う．

次に，目標ドメインにおける強識別器の学習 (図 5.3(b)) では，少数の目標ドメインの距離画像と教師信号から，距離画像の前処理，時空間特徴量の抽出を行い，事前ドメインで学習した弱識別器の中から目標ドメインの時空間特徴量に最適な強識別器を構築する．提案手法では，以上述べた事前ドメインのデータと少数の目標ドメインのデータを組み合わせた学習により，カメラ視点の変化に追従した強識別器を構築する．以下，図 5.3 の各処理の詳細について述べる．

5.3.1 距離画像の前処理

距離画像の前処理では，距離画像から人のシルエットを抽出し，抽出した人のシルエットを三面図に展開する．なお本研究において，三次元データを直交する三方向に投影変換した画像の組を三面図と呼ぶ．

距離画像中の人のシルエットの抽出は，距離画像を用いた背景差分で行う．距離画像の背景差分は，図 5.4 に示すように，事前に作成しておいた背景画像と入力された距離画像との距離値が閾値以上に変化した画素を物体領域として抽出する．距離画像の背景差分はシンプルな方法ながら，距離情報を利用することで，人のシルエットが高精度に抽出できる [39]．



(a) 背景の距離画像　(b) 入力の距離画像　(c) 人のシルエット

図 5.4: 距離画像の背景差分．

次に，図 5.5(a) に示す距離画像中のシルエットを，射影変換モデルを用いて図 5.5(b) に示す三次元点群に変換する．入力距離画像の座標を (u, v)，距離値を $d(u, v)$，距離画像センサの焦点距離を f とすると，三次元空間における空間座標 (x, y, z) は射影変換のモデルから式 (5.1) で計算できる．ここで，図 5.5(b) に示す三次元点群は，自由に視点を仮想的に移動することが可能である．

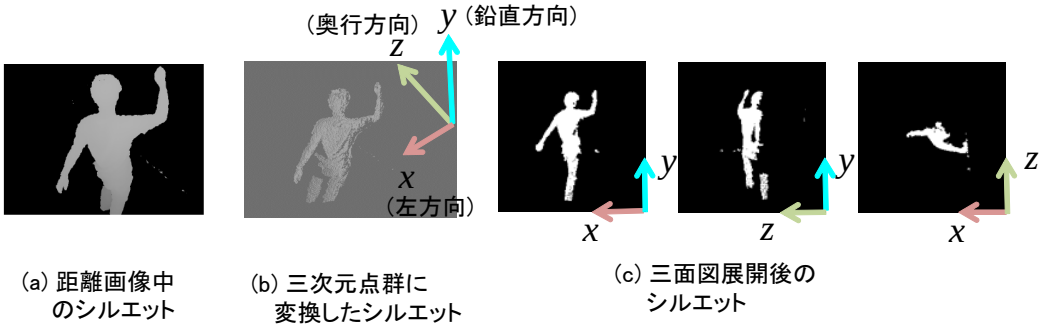


図 5.5: 距離画像の三面図展開.

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} ud(u,v)/f \\ vd(u,v)/f \\ d(u,v) \end{bmatrix} \tag{5.1}$$

次に，奥行，左，鉛直に対応した z，x，y 軸上の無限遠の仮想視点から，抽出したシルエット上の全ての画素を $x-y$ 平面，$z-y$ 平面，$x-z$ 平面にそれぞれ投影する．鉛直方向の y 軸は，距離画像センサが設置された俯角に応じて定まる．この三面図展開した投影画像には，物体の有無を示す二次元情報しかない．しかし，z，x，y 軸上の三方向から見た投影画像を合わせると三次元情報を反映することができる．また，この三面図は，仮想視点が無限遠にあるため，画面内の位置の変化による人の見えの変化を小さく抑えることができる．これは，図 5.1 の壁を殴る動作や壁を蹴る動作のように，カメラの視野内において人の位置の変化する状況で有効である．なお，入力距離画像上で遮蔽が発生した人体の部位については，三面図展開した際のシルエットに反映されない．

5.3.2　カメラ視点を変えたデータの生成

本節では，距離画像中の人の向きが変化するように，仮想的にカメラ視点を変化させてデータを生成する方法について述べる．距離画像中の人のシルエットは，式 (5.1) で三次元点群に変換した後，仮想的にカメラ視点を任意の位置および向きを変えることができる．本研究では，図 5.5 の距離画像から，鉛直方向 y 軸を回転軸として人の向きが変わるように，仮想的にカメラ視点を変化させる．ここでカメラ視点を変化させた後の奥行方向 z' 軸は視線方向にあたる．カメラ視点を変化させた後，三次元点群に変換した人のシルエットは，5.3.1 節と同様の手順で三面図に展開できる．視線方向を変化させたときの座標軸 (x', y, z') の例を図 5.6(a) に示し，各視線方向に応じた三面図展開後のシルエットを図 5.6(b) に示す．

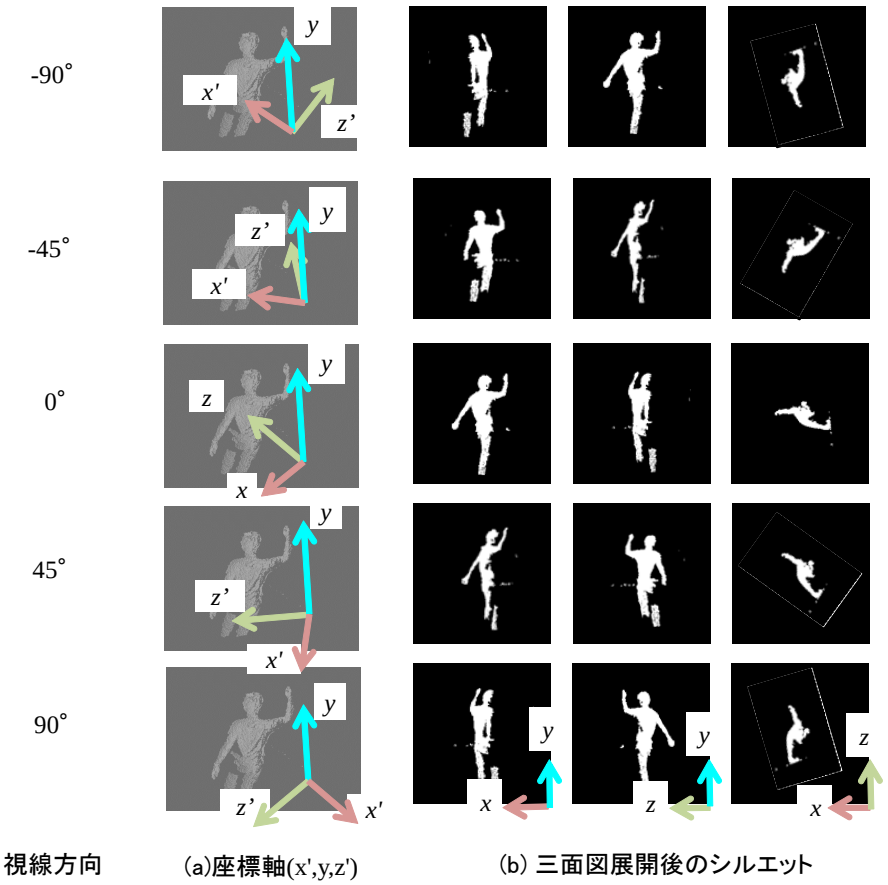


図 5.6: カメラ視点を変えたデータの生成の概要.

本節の方法で生成した三面図展開は，実際にカメラ視点を変えた距離画像から求めたときの三面図展開と比べると，自己遮蔽を無視した近似となっている．ここでの自己遮蔽とは，距離画像センサから見て人の遠方の部分が人の手前側の部分によって遮蔽されることを意味する．本節の方法で生成した三面図展開は，この自己遮蔽により人の手の動き等の細かな動きを捉えることはできないが，人体全身の大まかな特徴を捉えることができる．本研究で対象とする暴れ動作や転倒動作では，人体全身の動きを捉えて動作を認識するため，この自己遮蔽の影響は小さいと考えられる．

5.3.3 時空間特徴量の抽出

提案手法では，前処理した距離画像中の人のシルエットの見えと動きを捉える時空間特徴量として，Motion History Image(MHI) を利用する．MHI は，2.2.4 節で紹介した画像上の時間変化を濃淡で記録した特徴量である [28]．MHI の計算手順を図 5.7 に示す．まず MHI の各時刻の断片から輪郭の方向のヒストグラムを計算する．次にこの断片の面積で，ヒストグラムのビンの総数を正規化する．MHI 特徴量は，人体の動いた領域の形状と動きの方向，および動きの大きさを反映する．

MHI 特徴量は，三面図に展開した各平面において抽出する．ビン数が 18 のとき，時空間特徴量は 54 次元となる．これを時間方向に拡張して情報量を増やす [9]．時刻の数が 6 のとき，時空間特徴量は三平面合計で 324 次元である．

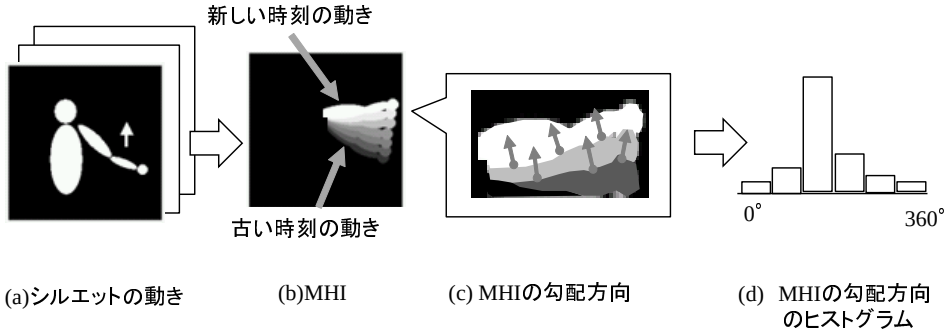


図 5.7: MHI を用いた時空間特徴量．

5.3.4 弱識別器の学習

事前ドメインにおける弱識別器は，カメラ視点に応じて生成した時空間特徴量毎に学習する．統計的学習法には，Random Forest[70] を用いる．Random Forest は複数の決定木から構成されたマルチクラスの識別が可能な識別器の一種であり，2001 年に Breiman により提案された．Random Forest は，ブートストラップ法により学習サンプルから複数の弱識別器を学習することによって，過学習を防ぎ高い汎化性能を持つ．高い汎化性能を持つこと，並列処理による高速な学習および認識が可能なこと，クラス識別，クラスタリング，回帰分析と多くの用途に適用できることから，コンピュータビジョンの分野でも人のスケルトンの認識 [39] や物体セグメンテーション [71] 等，広く用いられている．

Random Forest は，文献 [70] に記された学習アルゴリズムにより， 図 5.8 に示すような複数の決定木から構成される識別器を構築する．各決定木は，分岐ノード (Split Node) と末端ノード (Leaf Node) により構成され，分岐ノードを繰り返し作成し，一定の基準により分岐が不可能になった際に，末端ノードを作成することにより構築される．分岐ノードは，ある 1 次元の特徴量と閾値の大小関係により左右に分岐する．また，末端ノードは辿り着いた学習サンプルの各カテゴリの事後確率を保存する．

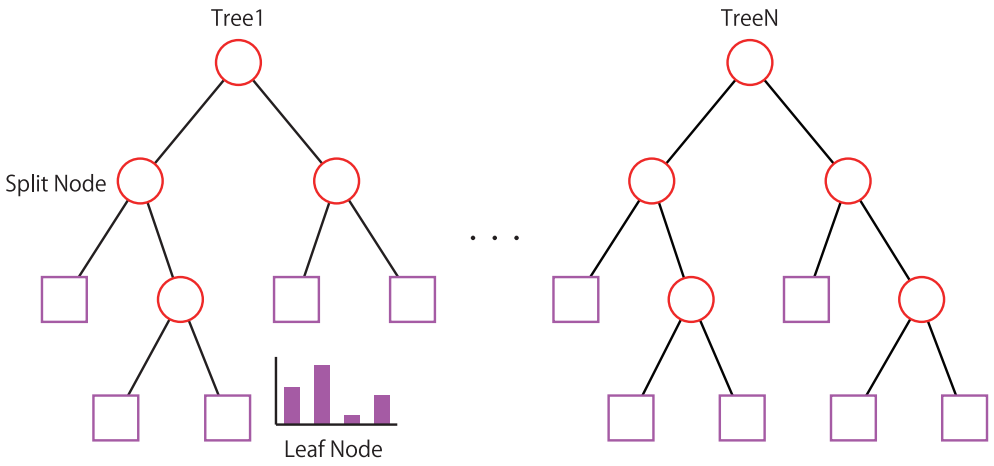


図 5.8: Random Forest の構造.

提案手法では，カメラ視点毎の時空間特徴量からそれぞれ Random Forest を学習し，構築した決定木それぞれを弱識別器候補とする．カメラ視点の種類を D，各 Random Forest の決定木の数を T とすると，弱識別器候補の総数は $D \times T$ となる．図 5.9(a) に弱識別器の学習の過程を示す．

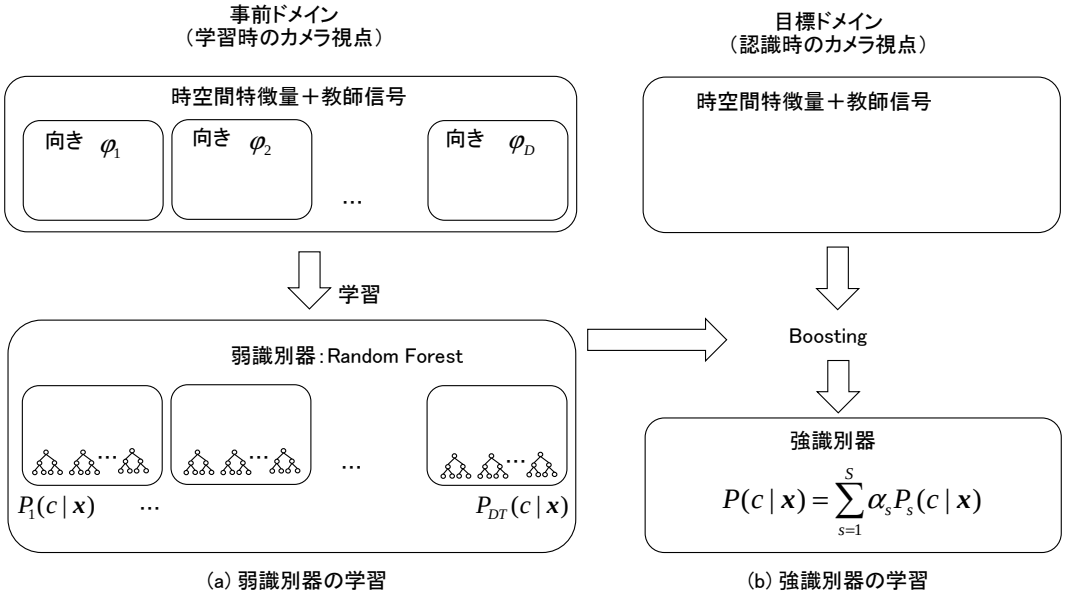


図 5.9: Random Forest を用いた弱識別器および強識別器の学習.

5.3.5 強識別器の学習

本節では，図 5.3(b) の目標ドメインにおける強識別器の学習方法について述べる．まず，与えら

れた目標ドメインの少数の距離画像のサンプルを対象に，距離画像を前処理し，時空間特徴量を抽出する．次に，抽出した時空間特徴量と，目標ドメインのサンプルの教師信号と，5.3.4 節で学習した弱識別器候補から，強識別器を学習する．以下，図 5.3(b) 中で既に説明した距離画像の前処理と時空間特徴量を除いて，Boosting により強識別器を学習する方法について述べる．

ここでは目標ドメインにおける少数の学習サンプルの時空間特徴量および教師信号と，カメラ視点に応じた Random Forest の決定木の集合 $\{P_i(c|x)_{i=1...TD}\}$ である弱識別器候補から，Boosting を導入した学習アルゴリズムを用いて式 (5.8) の強識別器を学習する．この学習アルゴリズムでは，Random Forest の決定木の集合の中から目標ドメインにおいて有効な決定木を選択して，有効な決定木群からなる強識別器を構築する．

Boosting の学習アルゴリズムでは，何らかの指標により弱識別器 $h(x)$ を逐次的に選択することで強識別器 $H(x)$ を学習する．ここで学習サンプルの重み w という概念を取り入れ，弱識別器 $h(x)$ を選択後に学習サンプルに対して重み付けする．この学習サンプルに対する重み w は，次の弱識別器 $h(x)$ を選択する際に，その学習サンプルをどれほど考慮するかを表す．正しく識別したサンプルに小さな重み，誤識別したサンプルに大きな重みを与える．これにより，次に弱識別器 $h(x)$ を選択する際に，学習サンプルの重みが大きいサンプルをより考慮するため，識別に失敗したサンプルを正しく識別することができる弱識別器 $h(x)$ が選択される．これを繰り返すことにより複数の弱識別器 $h(x)$ から構成される強識別器 $H(x)$ を学習する．以上の処理を，選択した弱識別器の正解率が識別クラスのカテゴリ数 C の逆数 $1/C$ 以上になる条件を満たす間，所定回繰り返す．**アルゴリズム 5.1** に強識別器 $H(x)$ を学習するアルゴリズムを示す．決定木の集合 $\{P_i(c|x)_{i=1...TD}\}$ 中には，仮想的にカメラ視点を変化させたデータから学習した決定木が含まれるため，この学習アルゴリズムにより目標ドメインに最適な決定木を選択することで，少数サンプルで強識別器の高精度化が可能になる．

アルゴリズム 5.1: Boosting を応用した強識別器の学習アルゴリズム.

1. 入力: J 個の学習サンプル $\{\boldsymbol{x}_1, y_1\} \ldots \{\boldsymbol{x}_J, y_J\}$ を用意する.
   $\boldsymbol{x}$ は特徴量, $y_i \in \{1, \ldots, C\}$ はクラスラベルである.

2. 初期化: 学習サンプルの重み w を初期化

$$w_{j,1} = 1/J \tag{5.2}$$

   弱識別器の重み $\{\alpha_1, \ldots, \alpha_S\}$ の初期値を 0 にする.

3. 学習:
   For $s = 1, \ldots, S$ //学習ラウンド

   · 全ての弱識別器からエラー率 ϵ_s が最小の弱識別器 $P_s(c|\boldsymbol{x})$ を選択

   · エラー率 ϵ_s の算出

$$\epsilon_s = \sum_{j \in \arg\max_c P_s(c|\boldsymbol{x}_j) \neq y_j}^{J} w_{j,s} \tag{5.3}$$

   · 弱識別器 P_s の重み α_s を算出

$$\alpha_s = \frac{1}{2} \log \left(\frac{(C-1)(1-\epsilon_s)}{\epsilon_s} \right) \tag{5.4}$$

   · 学習終了の判定

$$\text{if } \alpha_s < 0 \text{ then break} \tag{5.5}$$

   · 学習サンプルの重み w を更新

$$w'_{j,s+1} = \begin{cases} w_{j,s} \exp(+\alpha_s) & \text{if } \arg\max_c P_s(c|\boldsymbol{x}_j) \neq y_j \\ w_{j,s} \exp(-\alpha_s) & \text{otherwise} \end{cases} \tag{5.6}$$

   · 学習サンプルの重み w を正規化

$$w_{j,s+1} = \frac{w'_{j,s+1}}{\sum_{j=1}^{J} w'_{j,s+1}} \tag{5.7}$$

   End for

4. 出力: 強識別器

$$P(c|\boldsymbol{x}) = \sum_{s=1}^{S} \alpha_s P_s(c|\boldsymbol{x}) \tag{5.8}$$

5.4　生成型学習の学習結果を用いた動作認識

本節では，動作認識のフローを述べた後，フローの各ステップの詳細について述べる．図 5.10 に動作認識のフローを示す．図 5.10 では，まず認識時の距離画像を前処理して時空間特徴量を抽出する．次に，5.3.5 節で構築した強識別器を用いて動作カテゴリを識別する．最後に，動作の識別結果を時系列判定する．以下，学習時と共通である距離画像の前処理と時空間特徴量の抽出を除いた各処理について述べる．

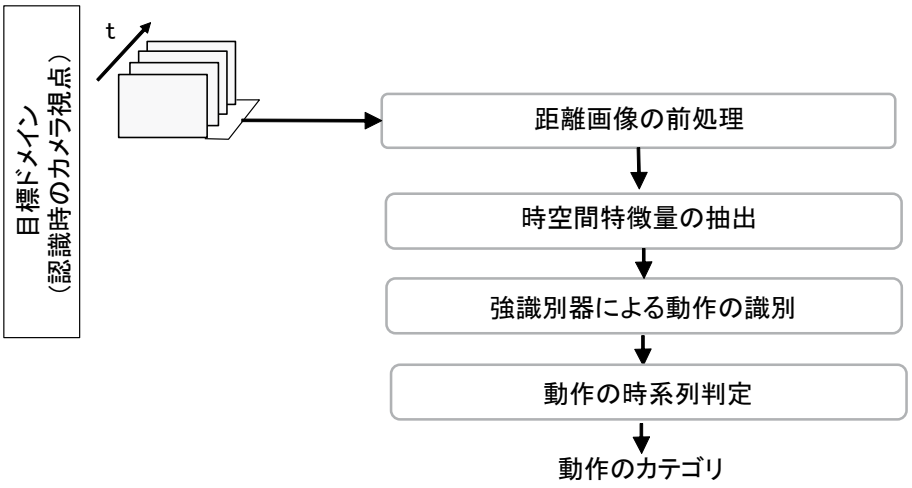


図 5.10: 提案手法の動作認識のフロー．

■ 強識別器による動作の識別

5.3.5 節で学習した式 (5.8) に示す強識別器を用いて，認識時の時空間特徴量から動作カテゴリを識別する方法について述べる．

識別時の概要を図 5.11 に示す．まず，認識時の時空間特徴量 $\boldsymbol{x}$ を各々の決定木に入力する．各決定木では，時空間特徴量 $\boldsymbol{x}$ に対して学習時に求めた分岐条件を用いて最上位のノードから末端方向のノードを辿り，対応する末端ノードを探す．そして，図 5.11 に示すように，各末端ノードにおける事後確率 $P_s(c|\boldsymbol{x})$ と重み係数 α_s の荷重和から，強識別器の事後確率 $P(c|\boldsymbol{x})$ を求める．最後に，式 (5.9) に示すように，$P(c|\boldsymbol{x})$ の事後確率が最大となる動作カテゴリを選択する．

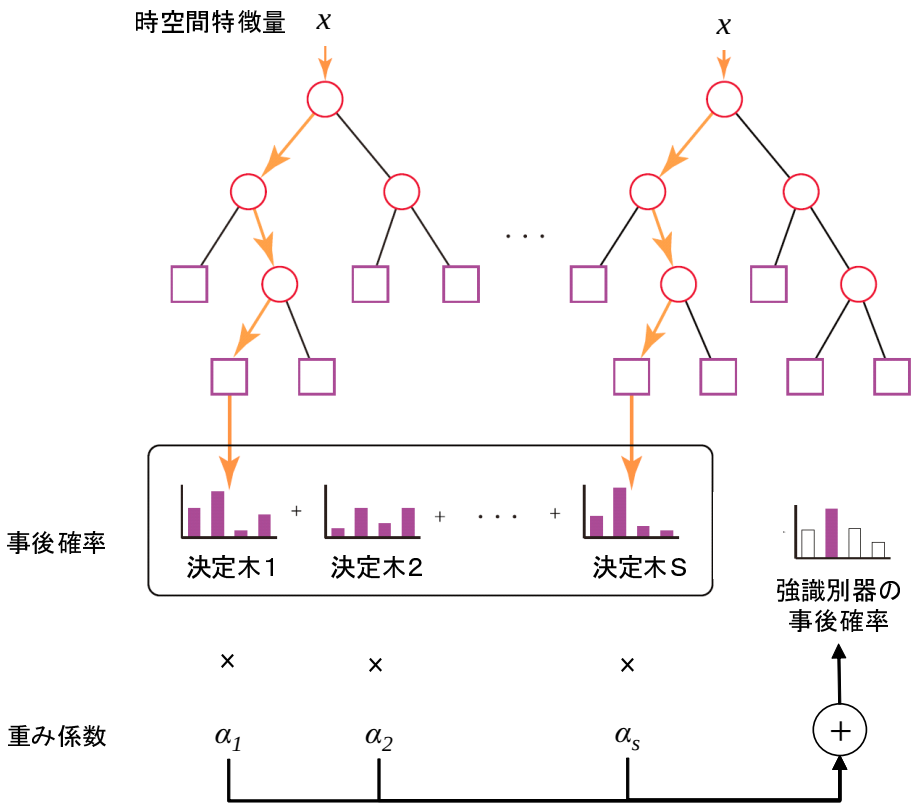


図 5.11: Random Forest を用いた強識別器による識別の概要.

$$\arg\max_c P(c|\boldsymbol{x}) \left(= \arg\max_c \sum_{s=1}^{S} \alpha_s P_s(c|\boldsymbol{x}) \right) \tag{5.9}$$

■ 動作の時系列判定

動作の時系列判定では，時空間の瞬時の乱れ等による誤認識を低減するために，直近 K フレームの時空間特徴量 $\{x_k\}_{k=\{1\ldots K\}}$ から式 (5.10) で求めた事後確率が最大の動作カテゴリ c を選択する．

$$\arg\max_c \prod_{k=1}^{K} P(c|\boldsymbol{x}_k) \tag{5.10}$$

5.5 実験条件

本節では，評価実験の実験条件について述べる．距離画像センサには，Microsoft 社の Kinect$^{\text{TM}}$ を用いた．この距離画像センサを室内の天井付近に取り付けて，斜め下を向けて撮影した．この距離

画像センサの視野内において，異なる方向を向いて所定の動作する人を撮影して評価データとした．

動作のデータは，図 5.12 に示す 8 種類のカテゴリとした．各動作の概要を述べると，“壁を殴る” では，人の正面の壁を殴る動作を続ける．“頭を掻く” では，頭を掻き続ける．“壁を蹴る” では，人の正面の壁を蹴る動作を続ける．“殴り合う” では，対面した 2 名が殴り合う動作を続ける．“歩く” では，入口から歩行して入り，入口の方向に振り向く．“鞄を奪う” では，対面した 2 名が鞄を掴んで引っ張り合う．“屈む” では，直立した状態から床面に屈み，屈み続ける．“背を伸ばす “では，直立した状態から腕を上に伸ばし下ろす動作を 1 回行う．“壁を殴る”，“壁を蹴る”，“殴り合う”，“鞄を奪う” は暴力行為，残りの動作は日常動作の例として選択した．学習用のカメラ視点のデータは，人の位置を入口付近とし，動作の向きを基本的に入口の方向 (図 5.1 中の 0°) とした．評価用のカメラ視点のデータは，人の位置を入口の横の壁付近とし，動作の向きを基本的に壁の方向 (図 5.1 中の-90°) とした．ただし，動きが激しい “殴り合う” および “鞄を奪う” では動作の途中で動作の向きがずれることがあった．また，“歩く”，“屈む”，“背を伸ばす” という一部の動作では，学習時も評価時も共通で動作の向きを入口の方向とした．動作を行う実験者は，学習用および評価用のデータで同一の 3 名とした．動作のデータのケース数は，学習用と評価用の 2 種類，動作の 8 種類，実験者の 3 名の組み合わせで，48 ケース取得した．動作のデータのフレーム数は，動作カテゴリと実験者の組み合わせで，学習用に計 1,235 フレーム，認識用に計 1,242 フレームを用いた．

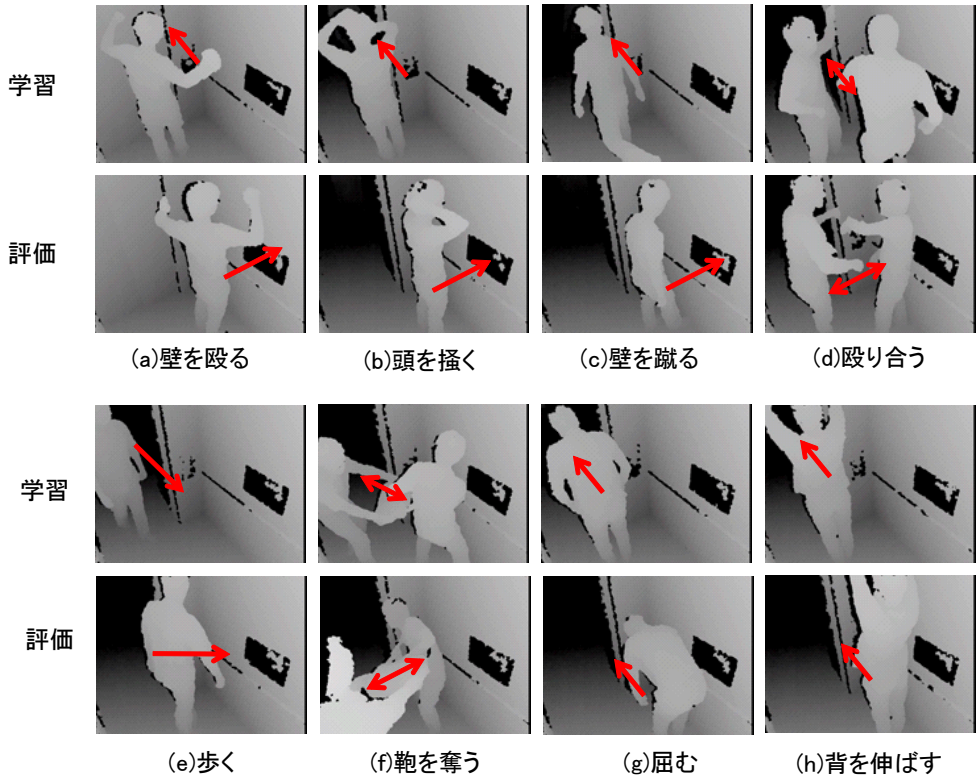


図 5.12: 評価実験のデータ.

提案手法におけるアルゴリズム中のパラメータは，予備実験から以下のように定めた．5.3.2 節のデータの生成において，人の向きは $-90°$ ～ $+90°$ の範囲で $5°$ 刻みに変化させた．この人の向きの生成の範囲は，図 5.12 に示すように学習時と評価時で人の向きがおおよそ $90°$ 異なることから定めた．5.3.4 節で述べた弱識別器の学習において，Random Forest の木の数 T は 100，木の深さ D は，特徴量の次元は 324(5.3.3 節参照)，特徴量選択数 F は 19，閾値選択回数 TH は 50，サブセット 1 つあたりの中のサンプル数は全サンプルの 75%とした．また，事前ドメインのデータは，学習用のデータ全てとした．5.3.5 節で述べた強識別器の学習において，弱識別器の選択数は 15 とした．5.4 節で述べた動作の時系列判定において，履歴長 K は 7 とした．

提案手法中における認識時のカメラ視点の学習サンプルのデータの量は，評価用のデータを母数として以下に述べるパラメータ τ により変化させた．評価用のデータの内，無作為に選んだ τ% のデータを目標ドメインの学習サンプルとした．そして，残る $100 - \tau$% を動作認識の性能評価用のデータとして用いた．

評価対象は，フレーム単位の認識結果とした．ある動作カテゴリに着目した時，真値がその動作カテゴリのフレーム数を N_T，認識結果がその動作カテゴリのフレーム数を N_R，認識結果が真値と一致したフレーム数を N_M とすると，Recall と Precision は式 (5.11) と式 (5.12) で計算される．また両者の代表値は，式 (5.13) に示す調和平均の F-Measure(F 値) で計算される．この F 値を各々の動作カテゴリで求めた後，代表値としてそれら F 値の平均値を求めて評価指標とした．

$$\text{Recall} = \frac{N_M}{N_T} \times 100 \tag{5.11}$$

$$\text{Precision} = \frac{N_M}{N_R} \times 100 \tag{5.12}$$

$$\text{F-Measure} = \left(\frac{\text{Recall}^{-1} + \text{Precision}^{-1}}{2} \right)^{-1} \tag{5.13}$$

5.6 実験結果

本節では，提案手法の実験結果について述べる．実験結果として，提案手法および以下に述べる 4 つの手法の全動作カテゴリ平均の F 値のグラフを図 5.13 に示す．図 5.13 のグラフの横軸は，評価用のデータ中で図 5.3 に示す目標ドメインのデータとして使う比率 τ である．

- 手法 1： 目標ドメインのデータから Random Forest を学習する手法.
- 手法 2 : 事前ドメインと目標ドメインのデータから Random Forest を学習する手法.
- 手法 3 : 事前ドメインと目標ドメインのデータおよびカメラ視点を変化させた生成データから Random Forest を学習する手法.

- 手法 4：原画像のシルエットから Random Forest を学習する手法.

手法 1 から 4 では，それぞれ所定のデータから距離画像の前処理と時空間特徴量の抽出を行った後に Random Forest を学習し，学習した Random Forest を用いて図 5.10 のフローで動作を認識する．手法 4 は，提案手法から 5.3.1 節の距離画像の前処理および 5.3.2 節のカメラ視点の視点を変えたデータの生成を省いたものである．この手法 4 では，距離画像の距離値を濃淡値としたカメラ画像として扱っている.

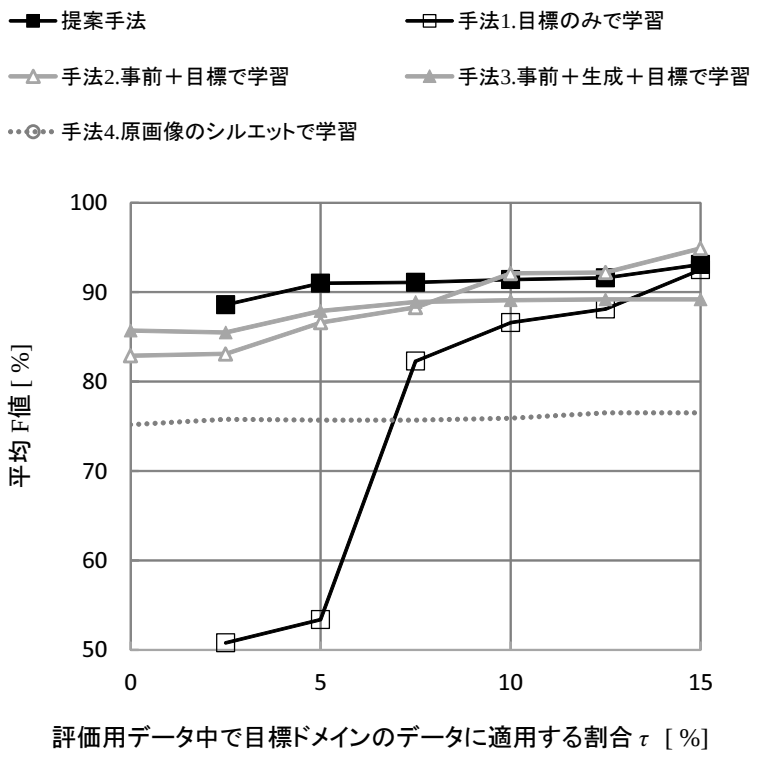


図 5.13: 実験結果のグラフ.

表 5.1: $\tau = 5\%$のときの F 値の比較.

手法	Recall[%]	Precision[%]	F 値 [%]
提案手法	90.7	91.3	91.0
手法 1	52.5	54.2	53.4
手法 2	84.4	88.9	86.6
手法 3	87.2	88.5	87.9
手法 4	75.6	75.9	75.7

図 5.13 の提案手法と手法 1 のグラフを比較すると，グラフの全域で前者は後者に優れ，特に τ が小さいほど大きな差があるが，τ が大きくなるにつれて両者の差は縮み，$\tau = 15\%$ではほぼ同じになる．提案手法と手法 2 のグラフを比較すると，前者は後者より $\tau < 10\%$の範囲で優れるが，$\tau \leq 10\%$の範囲では劣る．これは，手法 1 と手法 2 の目標ドメインを学習サンプルに加えることで人の向きの変化に追従するアプローチと比べて，提案手法が特に目標ドメインのデータが少ないときに効果が高いことを示す．

$\tau = 5\%$を代表値にとった時における各手法の F 値を表 5.1 に示す．表 5.1 において，提案手法は，手法 1，手法 2，手法 3，手法 4 と比べて，それぞれ 37.6%，4.4%，3.1%，15.3%優れている．ここで，提案手法と手法 3 の F 値の差 3.1%は Boosting による強識別器構築の効果 (5.3.5 節)，手法 3 と手法 2 の F 値の差 1.3%はカメラ視点を変えたデータの生成の効果 (5.3.2 節)，手法 2 と手法 4 の F 値の差 10.9%は三面図展開の前処理の効果 (5.3.1 節) を示している．

また，表 5.1 中の提案手法の F 値 91.0%と $\tau = 0\%$における手法 2(事前ドメインだけから Random Forest を学習) の F 値 82.9%を比較すると，前者は後者よりも 8.1%優れている．$\tau = 5\%$における手法 1(目標ドメインのデータだけから学習) の F 値が提案手法よりも 37.6%と大幅に劣ることを考慮すると，提案手法は目標ドメインの学習サンプルの量を有意に小さく抑えつつ，人の向きが 90° の変化する時の F 値を 8.1%改善したといえる．ここで，前述の構成要素の評価結果より，提案手法はカメラ視点を変えたデータの生成 (手法 3 と手法 2 の差)，Boosting による強識別器構築の効果 (提案手法と手法 3 の差) により，この F 値の改善を実現している．

5.7 まとめ

本章では，学習時と認識時の間におけるカメラ視点の変化が動作認識に与える影響を抑止する手法を提案し，評価実験により有効性を確認した．提案手法では，まず距離画像を人の向きの変化に伴う位置の変化に対して不変となる様に三面図展開で前処理する．次に，学習時のカメラ視点の距離画像のサンプルから，カメラ視点を変えた時の三面図を生成することで，生成範囲内のカメラ視点毎に応じた弱識別器を学習する．最後に，これら弱識別器と認識時と同じカメラ視点の少数のサンプルから最適な強識別器を構築する．

評価実験では，実験室内で実測した距離画像のデータを用いて，学習時と認識時とで人の向きが約 90° と大きく変わる変化する条件下において，暴れや歩行といった基礎的な動作の認識の精度が大きく改善することを確認した．評価用データの内 5%を目標ドメインの学習サンプルとした条件において提案手法の全動作カテゴリ平均の F 値は 91.0%であった．これは事前ドメインだけから Random Forest を学習する場合と比べて，目標ドメインの学習サンプルの量を有意に小さく抑えつつ F 値を 8.1%改善することができた．この F 値の改善は，カメラ視点を変えたデータの生成，Boosting による強識別器構築という提案手法中の各要素の効果に依る．

第6章

回帰推定による欠損距離画像の時空間特徴量補正

5章では，距離画像を用いた動作認識において，カメラ視点の変化に対して頑健な手法について述べた．しかしながら，この手法では動作する人のほぼ全身が映ることを前提としているため，人の近くに距離画像センサを設置するシーンでは，人体の一部が画角の外にはみ出て欠損すると対応できないという問題点がある．本章では，人の一部が画角の外にはみ出て欠損するシーンにおいて，この欠損の影響を抑止した動作認識の手法について述べる．

提案手法は，人の身体の一部が欠損したときの時空間特徴量と全身が映るときの時空間特徴量の相関性を利用した回帰分析を用いて，人の身体の一部が欠損したとき時空間特徴量を補正する．補正した時空間特徴量を用いて動作認識を実現することで，認識精度の改善を行う．

評価実験では，暴れや歩行等の基礎的な動作を対象とし，人の下半身が大きく欠損する条件において，動作認識精度が大きく改善することを確認した．

6.1　従来手法の課題

距離画像を用いた動作認識の従来手法において，人の一部が画角外にはみ出て欠損することが及ぼす影響について述べる．距離画像センサによる動作認識を屋内環境のモニタリングに活用する場合，一般的に画角の狭さが課題となる．これは距離画像センサの光学系が特殊なためであり，水平画角で比べると，監視カメラは 40∼110° 程度に対して，一般的な距離画像センサは 40∼70° 程度と狭い．

2.3 節で述べた距離画像を用いた動作認識の従来手法は，動作する人の全身が画角内に収まることを前提としている．しかし，距離画像センサの画角は狭いため，この前提は人の位置が画角中央の時は満たされても画角端部の時には満たされない．よって，人の位置が画角端部で人の全身が画角内に収まらない時は，時空間特徴量の一部が欠損し，動作カテゴリの判別が困難になる．この問題はジェスチャ動作 [33] や特定個所の動作 [35] を対象とする場合は，事前に人の位置を絞ることで回避できるが，セキュリティシステムが対象とする暴行や事故等の自発的な動作は，事前に人の位置を絞ることができないために大きな課題となる．

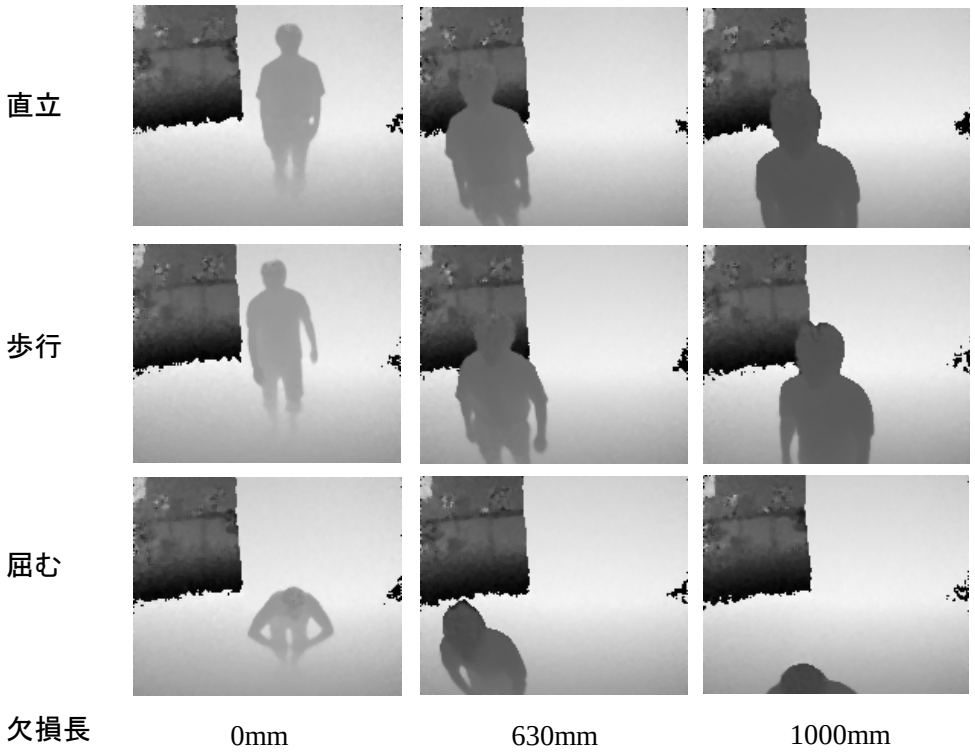


図 6.1: 欠損量毎の距離画像の例．

図 6.1 に距離画像中の人の下側が一部欠損するケースを示す．図 6.1 の第 2 列と第 3 列では，人の腿部あるいは腹部から下の部分が画角外にはみ出て欠損が生じていることがわかる．図 6.1 中の距離画像を認識対象とした際に，従来手法ではこのような欠損領域の局所的な見えや動きの情報

[37][33][34][35][36] や，スケルトンの間接情報 [40][42] が欠損することになり，時空間特徴量から動作カテゴリを判別することを困難にする．

図 6.1 中に示した腿部から下等の人体の下側の欠損は，距離画像センサが近い距離で人を撮影し，画角中の手前に人が映る時に発生する．また，距離画像センサが近い距離で人を撮影する際に，下側以外の個所が欠損することは無い．監視カメラが撮影対象とする人の近くにしばしば設置されることを考慮すると，下側の欠損に対処することは，監視カメラと同条件で設置した距離画像センサの動作認識の対象領域を広げることに大きく寄与できる．

6.2 提案手法のアプローチ

本章では，距離画像センサと動作する人の距離が近く人の下側が欠損した状況において，人の一部が欠損した時と全身が映る時の時空間特徴量の相関性を利用した回帰推定による時空間徴量の補正法を提案する．提案手法では，人の一部が欠損した時の時空間特徴量を補正して全身が映る時に近付けることで，欠損の影響の軽減するアプローチをとる．以下，6.3 節では，回帰推定による時空間特徴量の補正について述べる．6.4 節では，時空間特徴量の補正を用いた動作認識について述べる．6.5 節では，提案手法の評価実験の実験条件について述べる．6.6 節では，評価実験の実験結果について述べる．6.7 節では，まとめを述べる．

6.3 回帰推定による時空間特徴量の補正

距離画像を用いた動作認識で人の下側が一部が欠損した時を対象として，欠損時と全身が映る時の時空間特徴量の相関性を利用した回帰推定により時空間特徴量を補正する方法について述べる．手法の概要を図 6.2 に示す．まず，距離画像から 2.3 節で述べた人の見えや動きを反映した時空間特徴量を計算する．同時に，人の位置を求めて，位置に応じた人の欠損量を計算する．最後に，欠損量に応じた回帰係数を選択して，前記の欠損時の時空間特徴量から全身が映る時の時空間特徴量を回帰推定することで補正する．以下，各処理の詳細について述べる．

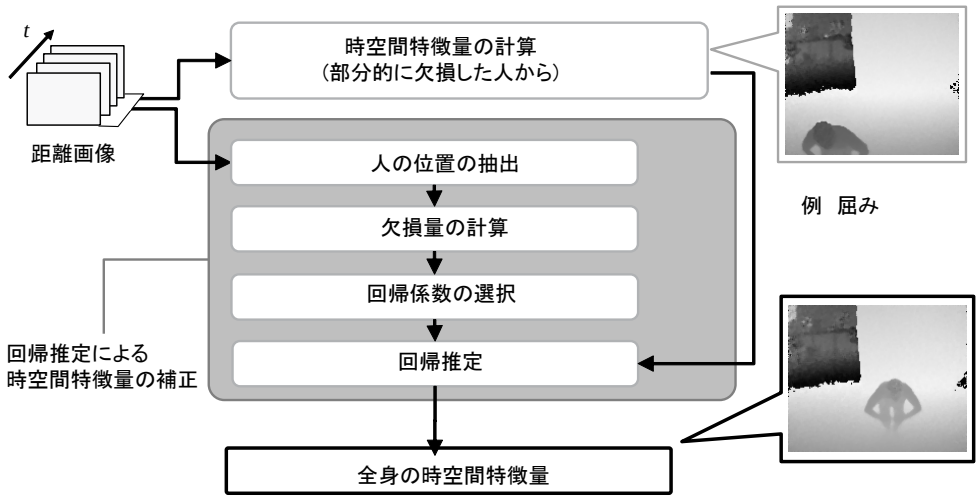


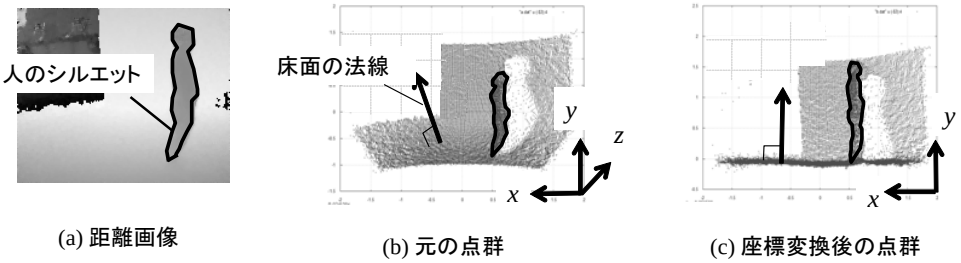
図 6.2: 回帰推定による時空間特徴量の補正の概要.

6.3.1 人の位置に応じた欠損量の計算

人の三次元空間における位置座標は，距離画像から人のシルエットを抽出して，シルエット内の画素の距離値から求める．欠損量は，求めた人の位置と距離画像センサの設置位置，設置角度，視野角を用いた幾何モデルから求めることができる．

■ 人の位置の抽出

距離画像中のシルエットは，5 章と同様に背景差分を用いて抽出する．距離画像の背景差分では，距離情報を用いてシルエットが高精度に抽出できる [39]．なお，背景は動作認識を行う前に求めておく．



(a) 距離画像　(b) 元の点群　(c) 座標変換後の点群

図 6.3: 距離画像の点群の座標変換の例.

人の位置の計算には，まず図 6.3 に示す様にシルエット中の各点を空間中の点群と捉えて，床面の

法線が鉛直となるように座標変換する [72]．次に，各点の床面 $x-z$ 上の重心を求めて人の位置とする．ここで x 軸は左方向，y 軸は上方向，z 軸は光軸方向である．

■ 欠損量の計算

欠損量は，距離画像センサから人までの床面上の距離 L と図 6.4 に示す距離画像の垂直視野角の幾何モデルから，式 (6.1) の B で計算する．なお B が 0 のときは，シルエットは画角内に収まり欠損は無い．

$$B = \max\big(0, Y_C - L/\tan(90-\theta-\omega/2)\big) \tag{6.1}$$

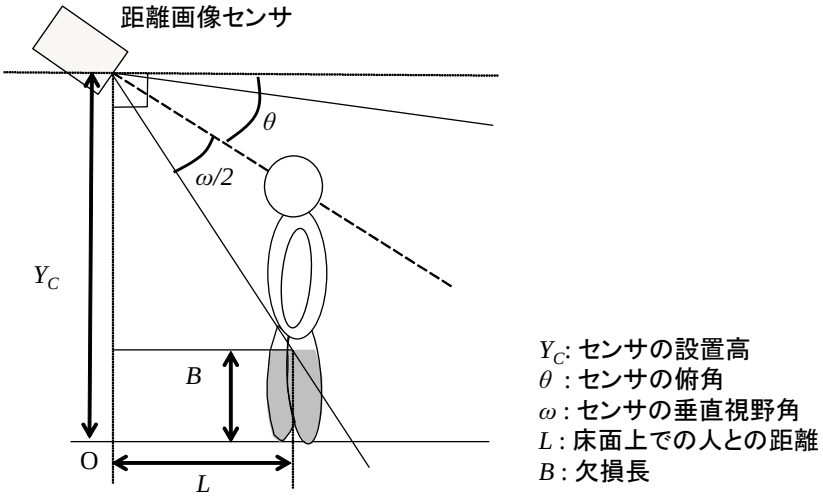


図 6.4: 人の位置と欠損量の図示．

6.3.2 欠損量に応じた時空間特徴量の回帰推定

時空間特徴量の回帰推定では，まず認識対象の距離画像から求めた欠損量に応じた応じて回帰係数 A_i および回帰分析の定数項 $\boldsymbol{c}_i$ を選択する．そして，欠損した距離画像から抽出した時空間特徴量を説明変数 $\boldsymbol{x}$ として，目的変数 $\boldsymbol{y}$ である全身が映る時の時空間特徴量を式 (6.2) で回帰推定する．

$$\hat{\boldsymbol{y}} = A_i\boldsymbol{x} + \boldsymbol{c}_i \tag{6.2}$$

以下，A_i および $\boldsymbol{c}_i$ の選択の方法を述べる．ここで事前に，欠損量は所定の間隔で N 種類のセット $\{B_1, B_2, \ldots B_N\}$ に離散化されている．また，$\{B_1, B_2, \ldots B_N\}$ の各々を欠損量とする距離画像から抽出した時空間特徴量を説明変数として，目的変数である全身が映る時の時空間特徴量を回帰推

定するための回帰係数のセット $\{A_1, A_2, \ldots A_N\}$ および定数項のセット $\{\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots \boldsymbol{c}_N\}$ が後節 6.3.3 の手順で計算済である．ここで，認識対象の距離画像から求めた欠損量が B の時，次式 (6.3) を用いて欠損量のセット中で B に最も近い B_i を求め，次に B_i と添え字が同じが同じ回帰係数 A_i および定数項 $\boldsymbol{c}_i$ を選択する．

$$\underset{i\in\{1,2,\ldots N\}}{\arg\min} \parallel B - B_i \parallel \tag{6.3}$$

6.3.3 欠損量に応じた回帰係数の計算

回帰係数のセット $\{A_1, A_2, \ldots A_N\}$ の各要素は，事前に学習サンプルの距離画像を用いて，欠損量のセット $\{B_1, B_2, \ldots B_N\}$ の全要素で事前に計算しておく．ここで欠損量 B_i の学習サンプルの距離画像は疑似的に，画角内に全身が映る学習サンプルの距離画像から，高さ B_i 以下の部分を除外して合成する．回帰係数 A_i は，回帰分析の係数項の計算式である式 (6.4) で計算する．式 (6.4) において，$S_{xx,i}$ は欠損量 B_i の学習サンプルの距離画像から求めた時空間特徴量の偏差平方和，$S_{xy,i}$ は全身が映る学習サンプルの距離画像ならびに欠損量 B_i の学習サンプルの距離画像から求めた時空間特徴量の偏差平方和である．定数項 $\boldsymbol{c}_i$ は回帰分析の定数項の計算式の式 (6.5) で計算する．式 (6.5) において，$\boldsymbol{\mu}_{x,i,}$ および $\boldsymbol{\mu}_y$ は，欠損量が B_i の時および全身が映る時の時空間特徴量の平均値である．

$$A_i = S_{xy,i}\, S_{xx,i}^{-1} \tag{6.4}$$

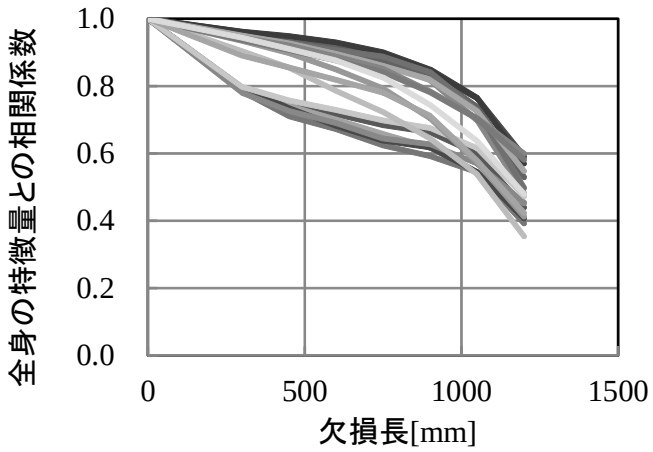
$$\boldsymbol{c}_i = \boldsymbol{\mu}_y - A_i \boldsymbol{\mu}_{x.i} \tag{6.5}$$

ここで式 (6.2) の $\hat{\boldsymbol{y}}$ は，説明変数 $\boldsymbol{x}$ に対する目的変数 $\boldsymbol{y}$ の変化を線形近似した時に，統計的に最小 2 乗で最適な推定値である．この回帰推定では，人の下側が一部欠損した時と全身が映る時とで時空間特徴量に相関性があることを前提にする．例えば，屈伸時に脚部が欠損する状況では，前者で捉える上体の上下の動きと，後者が捉える脚部を含めた全身の屈伸の上下の動きは同期しているので，この相関性が存在すると考えられる．

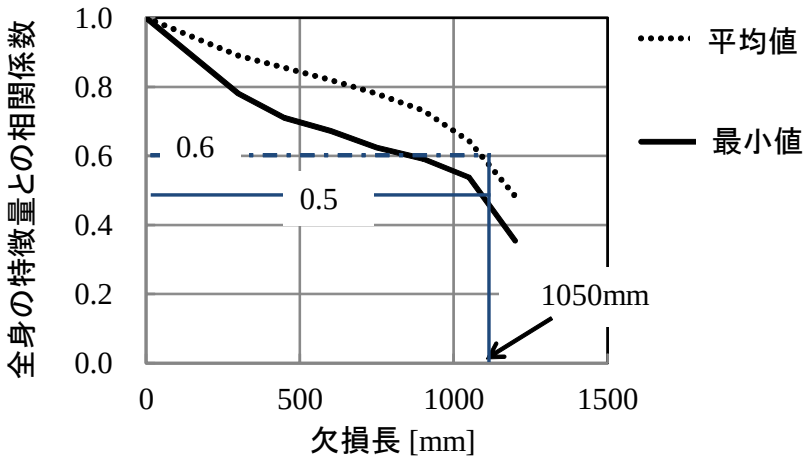
6.3.4 時空間特徴量の相関性の検証

人の下側が欠損した時と全身が映る時の時空間特徴量の相関性を検証した．図 6.5(a) に，後節 6.5 の実験データから求めた，所定の欠損量の時と全身が映る時の時空間特徴量の相関係数を示す．時空間特徴量は，5.3.3 節で述べた MHI を用いた時空間特徴量の各要素である．図 6.5(b) に代表値として，時空間特徴量の各次元の平均値および最低値を示す．相関係数の範囲は 0 から 1 であり，1 で一致，0 で無相関である．欠損量は 0 の時に全身が映り，0 から増加するに連れて床面から欠損する部分が増加する．図 6.5(a) では，どの要素の相関係数も，欠損量の増加に連れて単調減少するもの

の，減少の程度は緩やかである．これは，人の下側が一部欠損した時と全身が映る時の時空間特徴量に相関性があることを示す．図 6.5(b) は，1,050mm という大きな欠損量でも，最低 0.5，平均 0.6 の相関があることを示す．



(a)特徴量の要素間の相関係数



(b)相関係数の代表値

図 6.5: 欠損量毎の時空間特徴量の相関係数．

6.4　時空間特徴量の補正を用いた動作認識

時空間特徴量の補正法を含めた提案手法の動作認識の全体の概要を図 6.6 に示す．まず前処理で，距離画像から人のシルエットを抽出して三面図展開する．次に，前処理により求めた人のシルエッ

トの三面図から，見えと動きの時空間特徴量を抽出する．次に，6.3.3 節で述べた回帰推定で時空間特徴量を補正する．最後に，時空間特徴量から動作カテゴリを判別して，その結果を時系列判定でフィルタリングする．

以上述べたステップの内で，距離画像の前処理は 5.3.1 節で既に説明したものと共通である．また，時空間特徴量の抽出は，5.3.3 節で既に説明したものと，特徴量の次元を除けば共通である．5.3.3 節では MHI のビン数が 18 で時空間特徴量の次元が 324 に対し，本章では MHI のビン数を 6 とし，時空間特徴量の総次元を 108 とする．以下，動作カテゴリの判別および時系列判定について詳細を述べる．

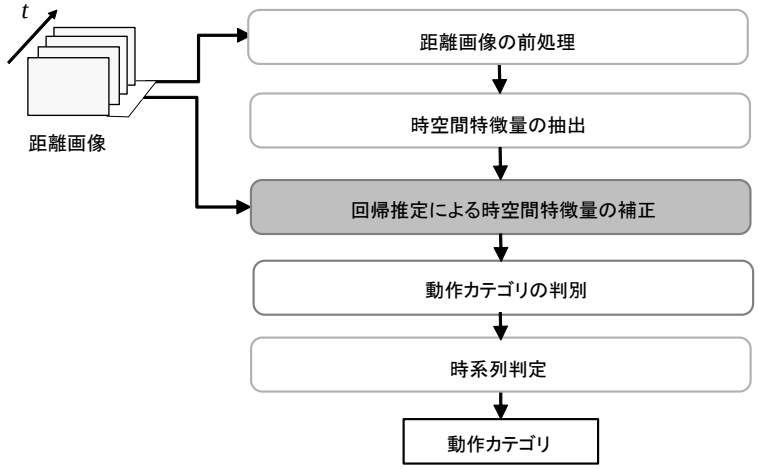


図 6.6: 時空間特徴量の補正を用いた動作認識の概要．

■ 動作カテゴリの判別

時空間特徴量から動作カテゴリを判別するには，Linear Discernment Analysis(LDA) による次元圧縮と kNN 法を用いる [9]．次元圧縮では，動作カテゴリの判別に寄与しない次元を削減して判別精度の向上を狙う．圧縮後の次元の選定は，LDA の固有値の寄与率 95%を基準とする．kNN 法は次元削減後の時空間特徴量に対して，式 (6.6) に示す代表ベクトルとの距離が最小の動作カテゴリに判別する．式 (6.6) で，$\boldsymbol{y}$ は時空間特徴量，$\boldsymbol{v}_{cm}$ は動作カテゴリ $c \in \{1, 2, \ldots C\}$ の代表ベクトル $\boldsymbol{v}_c = \{\boldsymbol{v}_{c_1}, \boldsymbol{v}_{c_2}, \ldots \boldsymbol{v}_{c_M}\}$ の m 番目の要素である．

$$\arg\min_{c} \parallel \boldsymbol{y} - \boldsymbol{v}_{cm} \parallel \tag{6.6}$$

この代表ベクトルは，事前に学習サンプルから LBG 法で求める [73]．学習サンプルには，画角内に全身が映るデータを適用する．

■ 動作カテゴリの時系列判定

時系列判定では，時空間特徴量の瞬時の乱れ等による誤認識の低減のために，認識結果の直近の履歴から式 (6.7) で求めた事後確率が最大の動作カテゴリ c を選択する．

$$\arg\max_{c} \prod_{k=1}^{K} P_{B_i}(c|\boldsymbol{v}_k) \tag{6.7}$$

式 (6.7) 中の K は履歴長で，予備実験より 18 とした．$\boldsymbol{v}_k$ は履歴中 k 番目に式 (6.6) で選択した代表ベクトルである．$P_{B_i}(c|\boldsymbol{v}_k)$ は，回帰係数の欠損量が B_i の時における $\boldsymbol{v}_k$ での動作カテゴリ c の事後確率である．この事後確率は，事前に代表ベクトルそれぞれで，欠損量のセット $\{B_1, B_2, \ldots B_N\}$ 全てにおいて式 (6.8) で計算しておく．式 (6.8) 中で $S_{c,i}$ は，欠損量 B_i の学習サンプル全てから式 (6.2) で補正値を求めた時に，代表ベクトル $\boldsymbol{v}$ を最近傍とする動作カテゴリ c の学習サンプルの数である．以上の手順により，この事後確率は欠損量毎の補正値に適した値となる．

$$P_{B_i}(c|\boldsymbol{v}_k) = \frac{S_{c,i}}{\sum_{j=1}^{C} S_{j,i}} \tag{6.8}$$

この時系列判定では更に，信頼度の低い認識結果の選別が可能である．認識結果の選別は，式 (6.7) 中で 1 番目と 2 番目に大きな動作カテゴリの事後確率 P_{1st} と P_{2nd} の比が閾値 T_p より大きな時には，事後確率の大小関係が偶発的とみなして認識結果を無効とする．

$$\frac{P_{2nd}}{P_{1st}} > T_p \tag{6.9}$$

6.5 実験条件

本節では，提案手法の評価実験の実験条件について述べる．距離画像センサには，一般的な ToF 方式のデバイスを用いた [31]．視野角は水平 41°，垂直 36° である．距離画像のサイズは 160 × 120[pixel] である．設置条件は，高さ 2.2m，俯角 25° とした．

動作カテゴリは図 6.7 に例を示す “屈伸”，“卒倒”，“揉合”，“振返”，“歩行”，“手振” の 6 つとした．“屈伸” では，中腰で脚部を数回曲げ伸ばす．“卒倒” では，直立から床に突伏す．“揉合” では，向き合う 2 名が腕を掴み合い揉み合う．“振返” では，直立から後方に振返る．“歩行” では，その場で足踏みをする．“手振” では，両腕を水平から真上まで数回振る．動作の内，“揉合” は暴力行為，“卒倒” は事故という異常動作の例，残りの動作は日常動作の例として選択した．また，距離画像センサに対する人の方向が動作認識に与える影響を考慮して，距離画像センサに対して正面と横向の 2 種とした．動作カテゴリの数は，動作 6 種と人の向き 2 種 (正面，横向) を組み合わせて 12 とした．動作中の人数は，“揉合” が 2 名，残りは 1 名である．動作のデータのケース数は，動作カテゴリ 12 個と，実験者 3 名と，位置 6 か所の組合せで 216 ケース取得した．位置の内訳は，図 6.1 の第 2，第 3 列に示す人の下側が一部欠損する 2 か所と，図 6.8 に示す人の全身が映る 4 か所である．動作のデー

タのフレーム数は，動作カテゴリと実験者の組み合わせで，各位置において順に 3,190，3,163，および 2,924，2,892，3,177, 2,892 フレームずつ取得した．どの位置でも 3,000 フレーム前後取得した

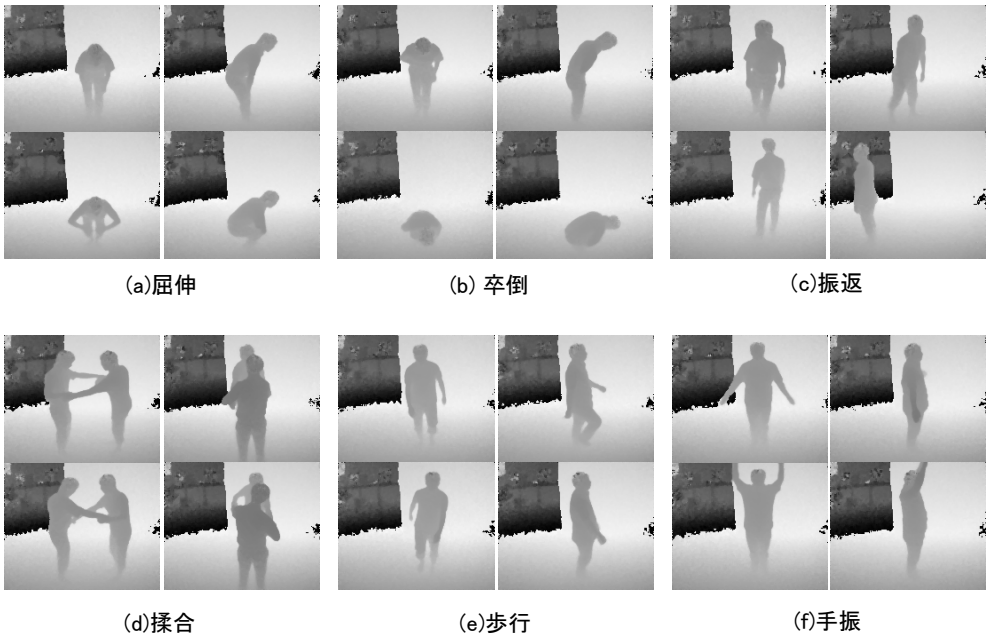


図 6.7: 動作データの例 (左列:正面，右列:横向)．

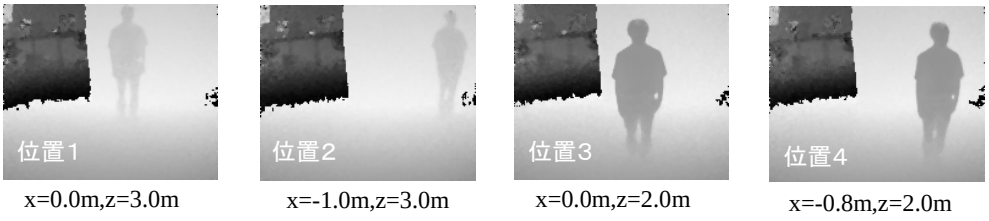


図 6.8: 位置毎の距離画像のサンプル．

時空間特徴量は MHI[28]，CHLAC[14]，ST-Patch[7] の 3 つとした．CHLAC はフレーム間差分を複数の局所パターンと照合した 251 次元の特徴量である．ST-Patch は，時間方向および空間方向の濃度勾配のモーメントによる 6 次元の特徴量である．どの時空間特徴量も，図 5.5 に示す三面図それぞれの成分のベクトル結合を時空間特徴量とする．ST-Patch は [9] 同様に時間方向に 6 フレーム累積した．

評価対象は，フレーム単位の認識結果とした．ある動作カテゴリに着目した時，真値がその動作カテゴリのフレーム数を N_T，認識結果がその動作カテゴリのフレーム数を N_R，認識結果が真値と一致したフレーム数を N_M とすると，Recall と Precision は式 (6.10) と式 (6.11) で計算される．ま

た両者の代表値は，式 (6.12) に示す調和平均の F 値 (F-Measure) で計算される．この F 値を各々の動作カテゴリで求めた後，代表値としてそれら F 値の平均値を求めて評価指標とした．

$$\text{Recall} = \frac{N_M}{N_T} \times 100 \tag{6.10}$$

$$\text{Precision} = \frac{N_M}{N_R} \times 100 \tag{6.11}$$

$$\text{F-Measure} = \left(\frac{\text{Recall}^{-1} + \text{Precision}^{-1}}{2} \right)^{-1} \tag{6.12}$$

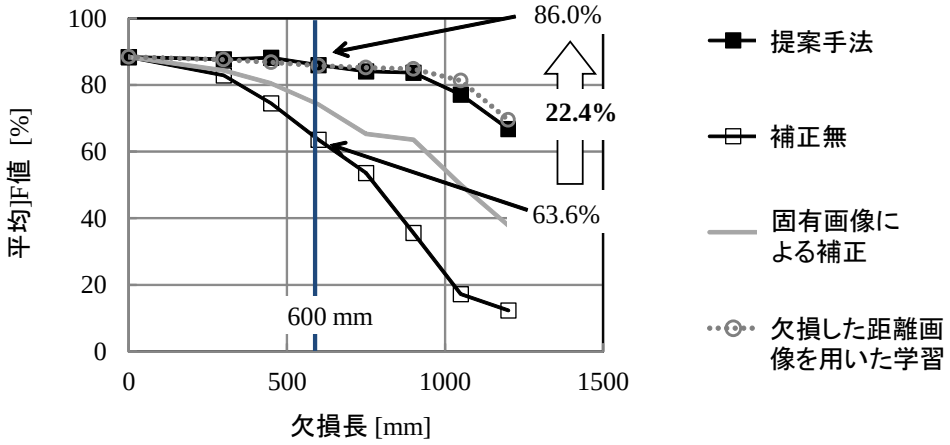
6.6 実験結果

提案手法の有効性の評価結果として，合成欠損データの評価結果，実欠損データの評価結果，欠損箇所を変えた時の基礎実験の評価結果，認識結果の選別機能の評価結果について述べる．

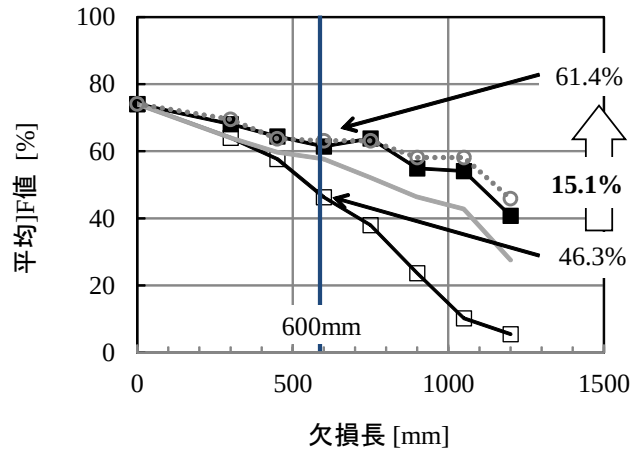
6.6.1 合成欠損データの評価結果

基礎評価として，以下の条件で提案手法の補正の有効性を評価した．評価対象のデータには，6.3.3 節と同じ手法で，欠損を合成した距離画像を用いた．6.3.1 節で述べた人の位置に応じた欠損量の計算を省き，本来 6.3.2 節の手法で選択すべき回帰係数の欠損量をデータの欠損量と同一にした．欠損の合成には，全身が映るデータから床面から欠損量の高さまでの点群を除外した．欠損量のセットは，300mm から 1,200mm まで 150mm 刻みとした．データは図 6.8 中で学習用を位置 2 と 3，評価用を位置 1 と 4 とした．学習用のデータは，欠損量毎の回帰係数および定数項を計算するのに用いた．なお，距離画像における位置毎の人の大きさや傾きの変化の影響は，図 5.5 に示す無限遠の仮想視点からの三面図展開による距離画像の前処理で低減させた．

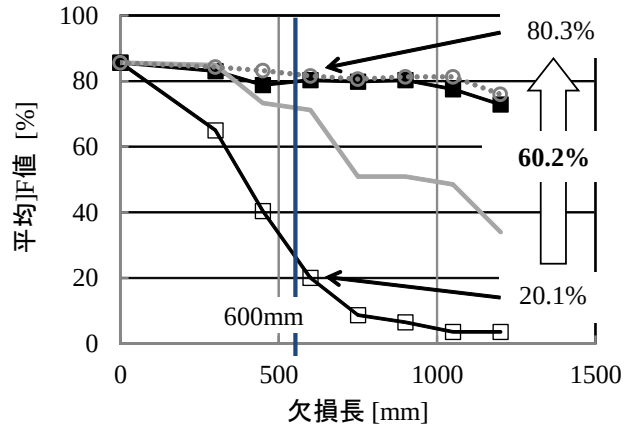
実験結果として，提案手法の全動作カテゴリ平均の F 値のグラフと，以下に述べる 3 つの手法のグラフを図 6.9 に示す．1 つ目は，時空間特徴量を補正しない手法である．2 つ目は，固有空間照合法において欠損部分を除いた部分画像から画像全体を復元する文献 [74] の方法で，距離画像 (図 5.5 中の $x-y$，$z-y$ 平面の投影画像) の欠損部を補間する手法である．補間した距離画像は，欠損量を 0mm として 6.4 節の手法で動作認識をした．3 つ目は，欠損量毎に動作カテゴリ判別の識別器を学習する手法である．事前に欠損量毎の距離画像から抽出した時空間特徴量から 6.4 節で述べた動作カテゴリ判別の識別器を学習し，認識の際には欠損を補間せずに動作カテゴリを判別する．



(a) MHI



(b) CHLAC



(c) ST-Patch

図 6.9: 合成欠損データの評価結果.

図 6.9 において提案手法で時空間特徴量を補正した時は，時空間特徴量を補正しない時と比べて欠損量 0mm から 1,200mm の全域で F 値が高いことから，提案手法の補正は有効といえる．欠損量 600mm の時に着目すると，F 値の改善は順に，22.4%，15.1%，60.2%である．成人男性の平均的な股下が 800mm であることから，この欠損量 600mm は脚部がほとんど見えない状況に相当する．なお，ST-Patch 特徴量の改善幅が特に大きいのは，他 2 つの時空間特徴量に無い見えの要素 (1 フレーム中の画像の特徴量) が要因と考えられる．見えの要素では，補正しない時は人の動きに依らず常に人が画角からはみ出た部分が欠損して小さくなるために，補正の効果が大きいと考える．また文献 [74] の手法と比較して提案手法は，ST-Patch の欠損量 300mm の条件では僅かに劣るものの残りの条件では F 値が上回ることから，全体では優れたと考えられる．

欠損量毎に識別器を学習する手法は，グラフが提案手法とほぼ重なるが僅かに上側の傾向にあり，時空間特徴量 3 種と欠損量 7 種の F 値の平均値は 1.6%上回った．ただし，この F 値の差は，提案手法と時空間特徴量を補正しない手法との F 値の同条件の平均値の差 33.2%と比べると軽微である．このように F 値の差が軽微である一方，提案手法は欠損量毎に識別器を学習する手法よりも，主記憶量が少ない点で優れている．提案手法および欠損量毎に識別器を学習する手法の主記憶量は式 (6.13) および式 (6.14) で計算できる．

$$NFDw + CM(D + C)w \tag{6.13}$$

$$NCM(D + C)w \tag{6.14}$$

式 (6.13) および式 (6.14) において，N は欠損量のセットの数，F は LDA 前の時空間特徴量の次元数，D は LDA 後の時空間特徴量の次元数，w は数値データ 1 つ当たりのバイト長，C は動作カテゴリの数，M は動作カテゴリ 1 つあたりの代表ベクトルの数である．MHI の時空間特徴量では N は 7，F は 108，L は 54，w は 8，C は 12，M は 100 であり，提案手法および欠損量毎に識別器を学習する手法の主記憶量は 0.9MByte および 4.4Mbyte であり，前者は後者よりも 80%少ない．この様な主記憶量の差は，特に低コストな組込計算機上で動作認識を行う時に重要となる．

6.6.2 実欠損データの評価結果

実際に欠損した距離画像において，提案手法の有効性を図 6.1 の第 2，3 列のデータで評価した．全評価データから式 (1) で計算した欠損量の中央値は図 6.1 の第 2 列が 630mm，第 3 列が 1,000mm である．評価データ以外の回帰係数の欠損量のセットおよび学習用データの条件は 6.6.1 節と同一とした．

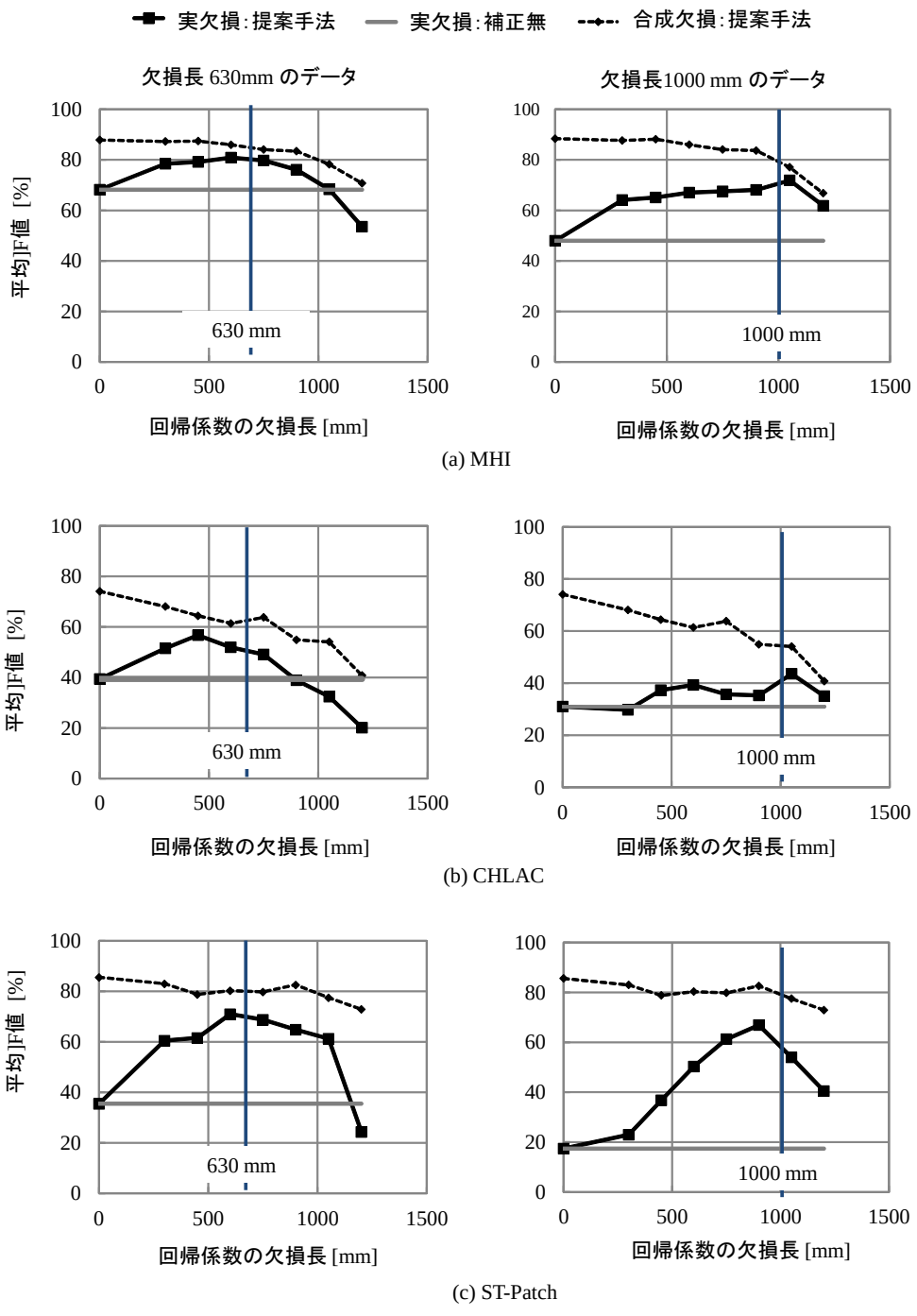


図 6.10: 実欠損データの評価結果.

評価結果のグラフを図 6.10 に示す．図 6.10 のグラフの横軸は 6.3.2 節で述べた回帰係数に対応した欠損量である．図 6.10 には，実欠損データを提案手法で補正した時と補正しない時のグラフを示す．また，比較として 6.6.1 節の合成欠損データを提案手法で補正したグラフを記す．まず，図 6.10

中で実欠損データを補正した時と補正しない時のグラフを比較すると，どの時空間特徴量でも実欠損量から ±150〜300mm の範囲の欠損量の回帰係数を適用した時に，補正した時は補正しない時から F 値が改善した．この回帰係数の欠損量の範囲は，提案手法における欠損量推定の許容誤差と考えられる．実欠損量に最も近い欠損量の回帰係数を選んだ時では，MHI，CHLAC，ST-Patch の順に，図 6.1 の第 2 列のデータでは欠損量 600mm の回帰係数の適用で 12.7%, 12.5%, 35.5%，図 6.1 の第 3 列のデータでは欠損量 1,050mm の回帰係数の適用で 23.9%, 12.6%, 36.6%の F 値が改善した．次に，図 6.10 中で実欠損と合成欠損を補正したグラフを比較すると，図 6.10 のどのグラフでも実欠損量の前後の欠損量の回帰係数を適用した時に両者は接近し，実欠損量と最も近い欠損量の回帰係数を適用した時で比較すると両者の F 値の差は最大で 10.5%に収まった．

6.6.3　欠損個所を変えた時の基礎実験の評価結果

提案手法中で欠損量の計算を除いて，時空間特徴量の補正が下側以外の欠損個所に適用可能かを，欠損個所を上側および右側としてして基礎評価した．上側の欠損は俯角が大きな距離画像センサの画角中で人が遠方のケース，右側の欠損は人が距離画像センサの画角左端に居るケースを想定する．評価データは 6.6.1 節と同様に合成欠損とした．上側の欠損は，実験者が腕を上に伸ばした時の高さ 2,150mm を欠損量 0mm，左側右側の欠損は実験者の重心から腕を水平に上げた時の重心から指先までの長さ 900mm を欠損量 0mm とした．時空間特徴量は MHI を対象とした．図 6.7 中の各カテゴリの動作の動き方を考えると，上側および右側の欠損時と全身の時空間特徴量は，下側の欠損時と同様に相関すると考えられる．

実験結果を図 6.11 に示す．上側および右側の欠損のどちらのグラフでも，全ての欠損量において提案手法の F 値は時空間特徴量を補正しない時より高く，それぞれ欠損量 600mm，900mm において F 値は 19.1%，21.0%改善した．この結果から，欠損個所の判別と，欠損個所に応じた欠損量が計算できれば，提案手法の時空間特徴量の補正は下側以外の欠損にも適用可能と考えられる．

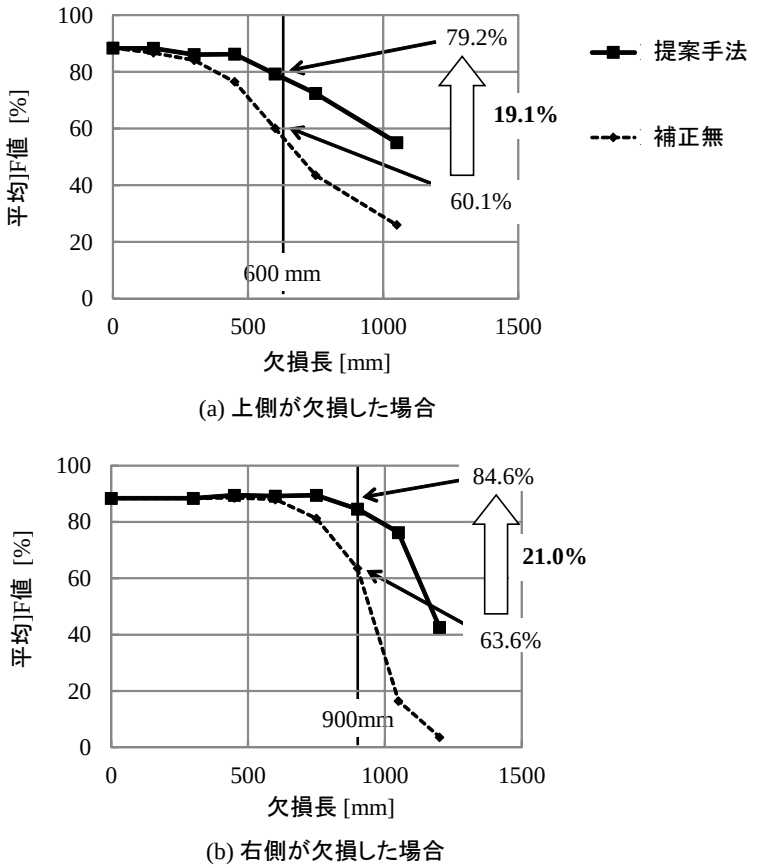


図 6.11: 欠損個所を変えた時の評価結果.

6.6.4 認識結果の選別機能の評価結果

式 (6.9) に示す認識結果の選別を動作カテゴリ毎の認識結果を対象に評価した．まず表 6.1 に，6.6.1 節の実験において認識結果を選別しない時の動作カテゴリ毎の F 値および Precision の変化量を示す．変化量は各々，欠損量 600mm と 0mm での F 値および Precision の差で計算した．表 6.1 中で F 値の変化量は特定カテゴリに集中し，平均の 2.4%減に対して，横向の “屈む” は 10.0%減，横向の “倒れ” は 19.9%減である．これらカテゴリの F 値の大幅な低下は，所定カテゴリ間で大幅に誤認識が増えたことが影響している．誤認識は，横向の “倒れ” のデータを横向の “屈む” もしくは正面の “倒れ” と間違うもので，下側の欠損により下半身の見える範囲が縮小することで，上半身の動きの類似した動作カテゴリ間で時空間特徴量が近付いたことが原因と推測する．表 6.1 中の Precision はこれら誤認識の増加により，横向の “屈む” で 16.5%減，正面の “倒れ” で 11.5%減と低下している．この様な特定カテゴリでの Precision の大幅な低下は，所定カテゴリの認識結果が重要なケース等では，認識結果全体の信頼性の低下に繋がるために改善が望まれる．

表 6.1: 欠損量 600mm 時の F 値と Precision の変化量.

動作	向き	F 値 [%]			Precision [%]		
		欠損長 [mm]		差分	欠損長 [mm]		差分
		0	600		0	600	
屈伸	正面	92.8	89.2	-3.6	87.9	87.1	-0.8
	横向	93.4	83.4	−10.0	88.0	71.5	−16.5
卒倒	正面	84.2	81.0	-3.2	87.8	76.3	−11.5
	横向	78.6	58.7	−19.9	88.5	92.3	3.8
振返	正面	82.7	84.1	1.4	81.9	87.4	5.5
	横向	89.6	93.0	3.4	84.9	94.7	9.8
揉合	正面	98.7	99.3	0.6	100.0	100	0.0
	横向	83.5	86.3	2.8	98.0	94.4	-3.6
歩行	正面	77.2	77.7	0.5	68.6	72.2	3.6
	横向	91.4	85.3	-6.1	86.8	85.3	-1.5
手振	正面	88.6	95.0	6.4	93.9	100.0	6.1
	横向	94.1	86.2	-7.9	88.8	83.0	-5.8
平均値 [%]		88.4	86.0	-2.4	—	—	—
最小値 [%]		—	—	—	—	—	−16.5

次に，式 (6.9) で認識結果を選別した結果を図 6.12 に示す．図 6.12(a) では，閾値の T_p を高くする程，F 値の変化量の全動作カテゴリの平均値が低下する一方で，Precision の変化量の全動作カテゴリの最低値は向上している．この F 値の低下は，正しい認識結果が無効と選別されることに依る Recall の低下が原因であり，F 値を評価指標とすることから最小限に留めることが望ましい．図 6.10(b) は T_p を介した両者のトレードオフの関係を示し，動作点の選択に利用できる．例えば Precision の変化量の最低値を 10%減より大きくする時は，Precision の変化量の最低値 8.2%減，F 値の変化量の平均値 4.6%減の動作点を選択する．

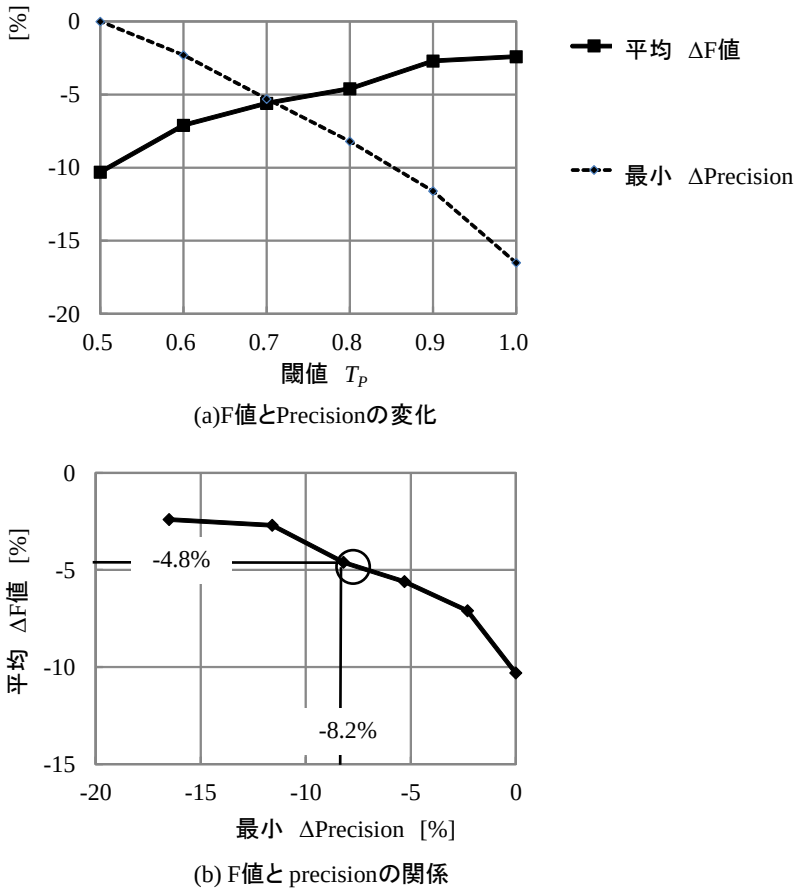


図 6.12: 欠損量 600mm 時の F 値と Precision の変化量.

6.7 まとめ

本章では，距離画像センサを用いた動作認識の対象領域を広げるために，動作する人と距離画像センサの距離が近い場合に画角外に人の下側が一部はみ出て欠損した時の時空間特徴量を回帰推定で補正する手法を提案した．また，実験室内の動作認識の基礎実験では，3 種類の時空間特徴量を対象に，提案手法の補正法で F 値が改善する効果を確認した．実験結果中で床面から 630mm まで上が実際に欠損したケースにおいて，時空間特徴量を補正する場合では補正しない場合よりも F 値を 12.5%以上改善した．この実験結果から，提案手法が画角の端部で人の下側が一部はみ出て欠損する影響を低減することで，距離画像センサの手前側における動作認識の対象領域を拡大できることを確認した．

加えて，欠損箇所を下側から右側および上側に変えた時の基礎実験の結果から，欠損箇所の判別と欠損量の推定が可能であれば，提案手法中の回帰推定による時空間特徴量の補正が下側に限定さ

れずに適用可能という見通しを示した．更に，事後確率による動作カテゴリの判別において，1 番目と 2 番目に事後確率が大きなカテゴリ間における事後確率の比を閾値と比較することで認識結果の有効無効を判別することにより，誤認識による Precisoin の低下を低減可能なことを示した．

第7章

結論と展望

本論文では，動画像中の見えと動きを捉える時空間特徴量と．背景や人の特徴を学習する統計的学習を用いることで，セキュリティ分野で課題となる環境条件の変化に頑健な動画像認識の手法について述べた．

7.1 結論

各章のまとめは以下の通りである．2 章では，本論文が提案する時空間特徴量の統計的学習を用いた動作認識のフレームワークについて述べた．このフレームワークは，動体検知，物体 (人) 検出，動作認識を構成要素とする．また，カメラ画像および距離画像を対象に，このフレームワークを構成する時空間特徴量をサーベイした．そして，カメラ画像と距離画像の時空間特徴量の比較を行い，それぞれの時空間特徴量の特長を明らかにした．

3 章では，ST-Patch 特徴量と呼ばれる時空間特徴量と混合正規分布を用いた背景モデルを用いて，カメラ画像中の動体を検知する手法を提案した．提案手法では，ST-Patch 特徴量により抽出した小領域単位の見えおよび動きの要素を用いることで，背景画像上の局所的ならびに大域的な変化を吸収して動体を検知する．評価実験では，屋外と屋内とエスカレータという 3 シーンにおいて，動きの情報を持たない見えの情報のみを用いた従来手法よりも優れた実験結果を取得した．この実験結果より，提案手法の有効性を示した．

4 章では，動体検知よりも詳細な情報を認識対象である人から取得するための手法として，ST-Patch 特徴量の統計的学習により人領域を検出し，人領域内の動きの非定常度を用いて突発的な異常動作を検出する手法について述べた．提案手法では，ST-Patch 特徴量から計算した非定常度に人領域検出の信頼度を重み付けることにより，画像上の動きが小さな異常動作でも検出がしやすくなる．評価実験では，背景中でエスカレータが定常的に動き続けるシーンにおいて，人の領域を検出し，転倒の動作を良好に検出できる実験結果を取得した．この実験結果から提案手法の有効性を示した．しかしながら，提案手法では，正面付近から撮影した人の全身の特徴量を用いて人領域を検出するために，動作を通じて人の姿勢の変化が大きな時や，カメラを斜めに設置したときには対応が難しいという課題点がある．また，非定常度を特徴量として動作を認識するために，認識対象とする動作の種類が転倒のような突発的な動作に限られるという課題点がある．

5 章では，距離画像センサと呼ばれる画像中の各画素の距離を計測するデバイスを用いることで，人の見えの変化に対して頑健かつより多くの種類の動作を認識対象とした手法について述べた．提

案手法では，距離画像を用いた動作認識において，カメラ視点の変化により人の向きが変化することが動作認識に及ぼす影響を抑止する．提案手法では，まず距離画像の学習サンプルを人の向きの変化に伴う位置の変化に対して不変となる様に三面図展開で前処理する．次に，学習時のカメラ視点の距離画像から，カメラ視点を網羅的に変化させた時の三面図を生成して，生成範囲内のカメラ視点毎に応じた弱識別器を学習する．そして，これら弱識別器と認識時と同じカメラ視点の少数のサンプルから最適な強識別器を構築する．これにより，学習サンプルの収集の手間を小さく抑えながら，カメラ視点が変化した時の人の動作を良好に認識することが可能となる．評価実験では，学習時と認識時で人の向きが約 90° 変化する条件下において，暴れや歩行等の基礎的な動作の認識精度が改善する実験結果を取得した．この実験結果から，提案手法の有効性を示した．しかしながら，提案手法では動作する人のほぼ全身から時空間特徴量が抽出できることを前提としているため，動作する人の近くに距離画像センサを設置するシーンでは，人の一部が画角外にはみ出て欠損するために対応困難である．

6 章では，距離画像センサの近くに人が来て，人の一部が画角の外にはみ出て欠損するシーンにも適用可能とするために，この欠損が動作認識に与える影響を抑止する手法を提案した．提案手法では，欠損したときの時空間特徴量と全身が映るときの時空間特徴量の相関性を利用した回帰推定により時空間特徴量を補正することで動作認識の精度を改善する．評価実験では，暴れや歩行といった基礎的な動作を対象とし，人の下半身が大きく欠損する条件において，動作認識の精度が大きく改善した．この実験結果から，提案手法の有効性を示した．

本論文における以上各章の研究成果は，2 章において述べた動体検知，物体検出，動作認識の手法から構成される動作認識のフレームワークにおいて以下の貢献を果たす．3 章および 4 章で述べた手法は，カメラ画像を用いた際のフレームワークにおいて，環境条件による背景の見えの変化に対する認識精度の改善に貢献する．3 章の手法は，背景画像が局所的かつ大域的に変化する際に，動体検知の精度改善に貢献する．4 章の手法は，背景画像が定常的に変化し続ける際に，高精度な物体検出を可能とし，転倒等の突発的な動作認識の精度改善に貢献する．5 章および 6 章で述べた手法は，距離画像を用いた際のフレームワークにおいて，環境条件による人の見えの変化に対する認識精度の改善に貢献する．5 章の手法は，カメラ視点の変化に伴い人の向きが変化する際に，動作認識の精度改善に貢献する．6 章の手法は，人の一部が画角外に欠損する際に，動作認識の精度改善に貢献する．

7.2 展望

本論文では，時間的に連続したカメラ画像ならびに距離画像中の見えと動きを捉える時空間特徴量と．背景や人の特徴を学習する統計的学習を用いることで，環境条件により発生する画像中の背景および人の見えの変化に頑健な動画像認識の手法について述べた．この動画像認識の手法については，カメラ画像ならびに距離画像を対象とした手法について述べた．カメラ画像中を対象とした手法では，背景に変化がある状況下において，動体の検知ならびに，人の転倒の動作の検出精度を向上させることができた．また，距離画像を対象とした手法では，カメラ視点の変化により人の向きが変化するとき，および人の身体の位置の一部が画角外に欠損するときの動作認識の精度を向上

させることができた．しかしながら，本論文中の動画像認識技術を実用的かつ汎用的な技術にするためには，幾つかの課題がある．

1 つ目の課題は，カメラ画像を認識対象とした手法の認識対象を拡充することである．本論文では，図 2.1 のフレームワーク中の動体検知，物体 (人) 検出，動作認識の一連の認識機能は実現できたが，動作認識は転倒のような突発的な動作に認識対象が限られた．カメラ画像からより多くの情報を引き出すためには，認識対象のバリエーションを拡充することが望ましい．背景および人の見えの変化に対する頑健さを維持しつつ，より多くの種類の動作を認識できるよう機能面で拡充することを今後の課題に挙げる．

2 つ目の課題は，距離画像を用いた動作認識において，より現実的な動作を認識対象とすることである．本論文で認識対象とした動作は，暴れの様なセキュリティ用途のニーズに直結した動作が含まれる一方，ほとんどの動作は歩行や手を振るなど基本的な動作に留まった．実環境の多種多様な箇所に距離画像センサを設置した場合には，多くの種類の人の動作が映ることが予期されるので，それら動作を認識して人の意図の理解や行動の予測に繋げることを今後の課題に挙げる．

3 つ目の課題は，図 2.1 のフレームワーク中に示した行動認識を実現することである．動体検知，物体検出，動作認識の各々の認識結果は，単独でもセキュリティシステムにとり有用な情報であるが，動画像をより高い次元で意味的に理解するには，各機能を有機的に連携させる必要がある．ここで，適用範囲の広さに優れたカメラ画像を対象とした認識機能と，環境条件の変化に対する頑健さに優れた距離画像を対象とした認識機能を連携させれば，両者の特長を生かした行動認識を実現することができる．この様な優れた行動認識を実現することを，今後の課題に挙げる．

謝　辞

本研究の遂行にあたり，常日頃ご指導を賜りました中部大学工学部情報工学科 藤吉弘亘教授に深く感謝の意を表します．本論文をまとめるにあたり，有益なご討論，ご助言を賜りました北海道大学大学院情報科学研究科 金子俊一教授，中部大学工学部情報工学科 岩堀祐之教授，平田豊教授に謹んで感謝いたします．本研究で用いるアルゴリズムの研究開発およびデータの撮影にて多大な協力を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) 数井誠人氏，株式会社日立産業制御ソリューションズ (当時，中部大学工学部情報工学科) 村井泰裕氏に厚く御礼申し上げます．本研究における国内外発表および学術論文投稿に支援を頂いた株式会社日立製作所 三好雅則氏に謹んで深謝します．最後に，本研究に協力して頂いた機械知覚&ロボティクスグループの山下隆義講師，山内悠嗣助手と皆様に心より感謝致します．

参考文献

[1] I. Laptev: “On Space-Time Interest Points”, *Int. J. Computer Vision*, Vol. 64, No. 2, pp. 107–123, 2005.

[2] C. Harris and M. Stephens: “A combined corner and edge detector”, in proceedings of *In Proc. of Fourth Alvey Vision Conference*, pp. 147–151, 1988.

[3] I. Laptev, M. Marszalek, C. Schmid, B. Rozenfeld, I. Rennes, I. I. Grenoble, and L. Ljk: “Learning realistic human actions from movies”, in proceedings of *IEEE Computer Vision and Pattern Recognition*, pp. 1–8, 2008.

[4] N. Dalal and B. Triggs: “Histograms of Oriented Gradients for Human Detection”, in proceedings of *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1 - Volume 01*, CVPR '05, pp. 886–893, Washington, DC, USA, 2005. IEEE Computer Society.

[5] R. Chaudhry, A. Ravich, G. Hager, and R. Vidal: “Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions”, in proceedings of *in In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1932–1939, 2009.

[6] L. Fei-fei and P. Perona: “A bayesian hierarchical model for learning natural scene categories”, in proceedings of *In CVPR*, pp. 524–531, 2005.

[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 Transactions on Pattern Analysis and Machine Intelligence*, Vol. 29, No. 11, pp. 2045–2056, 2007.

[8] Y. Ke, R. Sukthankar, and M. Hervert: “Event Detection in Crowded Videos”, in proceedings of *IEEE International Conference on Computer Vision*, pp. 8–15, 2007.

[9] M. Kazui, M. Miyoshi, and S. Muramatsu: “Incoherent Motion Detection using a Time-series Gram Matrix Feature”, in proceedings of *International Conference on Pattern Recognition*, pp. 1–5, 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, 第 1 巻, pp. 21–31, Jly 2008.

[11] T. Furukawa, H. Fujihoshi, and A. Nomura: “A Method for Estimating Cut-Edit Points in Personal Videos”, in proceedings of *IEEE International Conference on Multimedia & Expo*, 2010.

[12] Y. Murai and H. Fujiyoshi: “Shot Boundary Detection Using Co-occurrence of Global Motion in Video Stream”, in proceedings of *International Conference on Pattern Recognition*, 2008.

[13] 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.

[14] T. Kobayashi and N. Otsu: “Action and Simultaneous Multiple-Person Identification Using Cubic Higher-Order Local Auto-Correlation”, in proceedings of *International Conference on Pattern Recognition*, pp. 741–744, 2004.

[15] T. Kurita, N. Otsu, and T. Sato: “A Face Recognition Method Using Higher Order Local Auto-correlation and Multivariate Analysis”, *In Proc. International Conference on Pattern Recognition*, pp. 213–216, 1992.

[16] 岩田健司, 佐藤雄隆, 小林匠, 依田育士, 坂上勝彦, 大津展之: “CHLAC による映像サーベイランスのためのビジュアルフレームワーク”, 第 13 回画像センシングシンポジウム, pp. 1–7, 2007.

[17] 中村元美, 河野賢治, 三好雅則: “昇降機の安全・安心を提供する遠隔監視システム”, 日立評論, pp. 742–745, 2008.

[18] B. D. Lucas and T. Kanade: “An Iterative Image Registration Technique with an Application to Stereo Vision”, in proceedings of *Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2*, IJCAI’81, pp. 674–679, San Francisco, CA, USA, 1981. Morgan Kaufmann Publishers Inc.

[19] 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, pp. 273–278, 2004.

[20] H. Wang, A. Kläser, C. Schmid, and C.-L. Liu: “Dense trajectories and motion boundary descriptors for action recognition”, *International Journal of Computer Vision*, Vol. 103, No. 1, pp. 60–79, May 2013.

[21] H. Wang, A. Klaser, C. Schmid, and C.-L. Liu: “Action Recognition by Dense Trajectories”, in proceedings of *CVPR*, pp. 3169–3176, Jly 2011.

[22] M. Marszalek, I. Laptev, and C. Schmid: “Actions in context”, in proceedings of *CVPR*, pp. 2929–2936, Jly 2009.

[23] J. Liu, J. Luo, and M. Shah: “Recognizing realistic actions from videos in the wild”, in proceedings of *CVPR*, pp. 1996–2003, Jly 2009.

[24] 都築勇司, 藤吉弘亘, 金出武雄: “SIFT 特徴量に基づく Mean-Shift 探索による特徴点追跡”, 情報処理学会論文誌. コンピュータビジョンとイメージメディア, Vol. 49, No. 6, pp. 35–45, mar 2008.

[25] D. G. Lowe: “Distinctive Image Features from Scale-Invariant Keypoints”, *Int. J. Comput. Vision*, Vol. 60, No. 2, pp. 91–110, November 2004.

[26] D. Comaniciu, V. Ramesh, and P. Meer: “Real-Time Tracking of Non-Rigid Objects using Mean Shift”, pp. 142–149, 2000.

[27] 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, pp. 1–9, 2008.

[28] G. Bradski and J. Davis: “Motion Segmentation and Pose Recognition with Motion History Gradients”, in proceedings of *IEEE Winter Conference on Applications of Computer Vision*, pp. 238–244, 2000.

[29] M.-K. Hu: “Visual Pattern Recognition by Moment Invariants”, *IRE Transactions on Information Theory*, pp. 179–187, 1962.

[30] D. DeMenthon: “Spatio-Temporal Segmentation of Video by Hierarchical Mean Shift Analysis”, *In Statistical Methods in Video Processing Workshop*, pp. 1–20, 2002.

[31] MesaImaging 社. “http://www.mesa-imaging.ch/”.

[32] PrimeSense 社. “http://www.primesense.com/”.

[33] M. Holte, T. Moeslund, and P. Fihl: “Fusion of range and intensity information for view invariant gesture recognition”, in proceedings of *Workshop on Time-of-Flight based Computer Vision*, pp. 1–7, 2008.

[34] W. Li, Z. Zhan, and Z. Liu: “Action Recognition Based on A Bag Of 3D Points”, in proceedings of *Workshop on CVPR for Human Communicative Behavior Analysis (in conjunction with CVPR 2010)*, pp. 9–14, 2010.

[35] 池村翔, 藤吉弘亘: “時空間情報と距離情報を用いた Joint Boosting による動作識別”, 電気学会論文誌 C, Vol. C, No. 9, pp. 1554–1560, 2010.

[36] B. Ni, G. Wang, and P. Moulin: “RGBD-HuDaAct: A Color-Depth Video Database for Human Daily Activity Recognition”, in proceedings of *Workshop on Consumer Depth Cameras for Computer Vision (in conjunction with ICCV 2011)*, pp. 1147–1153, 2011.

[37] L. Schwarz, D. Mateus, and N. Navab: “Manifold learning for ToF-based human body tracking and activity recognition”, in proceedings of *British Machine Vision Conference*, pp. 80.1–80.11, 2010.

[38] O. Oreifej and Z. Liu: “HON4D: Histogram of oriented 4D normals for activity recognition from depth sequences”, in proceedings of *In CVPR*, 2013.

[39] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, , and A. Blake: “Real-Time Human Pose Recognition in Parts from a Single Depth Image”, in proceedings of *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1297–1304, 2011.

[40] S. Masood, C. Ellis, A. Nagaraja, M. Tappen, J. LaViola, and R. Sukthankar: “Measuring and Reducing Observational Latency when Recognizing Actions”, in proceedings of *Workshop on Human Computer Interaction (in conjunction with ICCV 2011)*, pp. 422–429, 2011.

[41] L.Xia, C. Chen, and J.Aggarwa: “View Invariant Human Action Recognition Using Histograms of 3D Joints”, pp. 20–27, 2012.

[42] J. Wang, Z. Liu, Y. Wu, and J. Yuan: “Mining Actionlet Ensemble for Action Recognition with Depth Cameras”, in proceedings of *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1290–1297, 2012.

[43] M. Gonen and E. Alpaydin: “Multiple Kernel Learning Algorithms”, *J. Mach. Learn. Res.*, Vol. 12, pp. 2211–2268, July 2011.

[44] H. Koppula, R. Gupta, and A. Saxena: “Learning Human Activities and Object Affordances from RGB-D Videos”, *IJRR*, Vol. 32, No. 8, pp. 951–970, 2013.

[45] P. Perez: “Markov random fields and images”, *CWI Quarterly*, pp. 413–437, 1998.

[46] A. Elgammal, D. Harwood, and L. Davis: “Nonparametric Model for Background Subtraction”, *European Conference on Computer Vision*, pp. 751–767, 2000.

[47] 島田敬士, 有田大作, 谷口倫一郎: “適応的な分布数の増減法を利用した混合ガウス分布による高速な動的背景モデル構築”, 電子情報通信学会論文誌, Vol. J90-D, No. 9, pp. 2606–2614, 2007.

[48] 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”, 電子情報通信学会論文誌, Vol. J86-DII, pp. 616–624, 2003.

[49] 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”, 電子情報通信学会論文誌, Vol. J79-DII, No. 4, pp. 568–576, 1996.

[50] M. Heikkila, M. Pietikainen, and J. Heikkila: “A Texture-based Method for Detecting Moving Objects”, in proceedings of *BRITISH MACHINE VISION CONFERENCE*, pp. 657–662, 2004.

[51] T. Chen, C. Chen, C. Huang, and Y. Hung: “Efficient hierarchical method for background subtraction”, *Pattern Recognition*, Vol. 40, pp. 2706–2715, 2007.

[52] 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, pp. 127–134, 2009.

[53] R. Vemulapalli and R. Aravind: “Spatio-temporal nonparametric background modeling and subtraction”, in proceedings of *IEEE International Workshop on Visual Surveillance (in conjunction with ICCV 2009)*, pp. 1145–1152, 2009.

[54] Y. Zhao, H. Gong, L. Lin, and Y. Jia: “Spatio-temporal Patches for Night Background Modeling by Subspace Learning”, in proceedings of *IEEE International Conference on Pattern Recognition*, pp. 1–4, 2008.

[55] A. Shimada and R. Taniguchi: “Hybrid Background Model using Spatial-Temporal LBP”, in proceedings of *IEEE International Conference on Advanced Video and Signal based Surveillance*, pp. 19–24, 2014.

[56] A. P. Dempster, N. M. Laird, and D. B. Rubin: “Maximum likelihood from incomplete data via the EM algorithm”, *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, Vol. 39, No. 1, pp. 1–38, 1977.

[57] T. Ellis and M. Xu: “Object Detection and Tracking in an Open and Dynamic World”, 2001.

[58] 公開映像データ. “http://limu.ait.kyushu-u.ac.jp/dataset/”.

[59] B. Scholkopf, J. Platt, J. Shawe-Taylor, A. Smola, and R. Williamson: “Estimating the Support of a High-Dimensional Distribution”, in proceedings of *Neural Comput*, Vol. 13, pp. 1443–1471, 2001.

[60] H. Zhong, J. Shi, and M. Visontai: “Detecting Unusual Activity in Video”, in proceedings of *Computer Vision and Pattern*, 第 2 巻, pp. 819–826, 2004.

[61] 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. 15, pp. 43–50, 2005.

[62] P. Viola and M. Jones: “Robust Real-Time Face Detection”, *Journal of Computer Vision*, Vol. 57, No. 2, pp. 137–154, 2004.

[63] B. Wu, H. Ai, C. Huang, , and S. Lao: “Improved Boosting Algorithms Using Confidence-rated Predictions”, *Machine Learning*, No. 37, pp. 297–336, 1999.

[64] 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.

[65] 弓場竜, 三好雅則, 藤吉弘亘, 村井泰裕: “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”, 電子情報通信学会論文誌, Vol. Vol.J94-D, No. 7, pp. 1101–1112, Jly 2011.

[66] D. Weinland, R. Ronfard, and E. Boyer: “Free viewpoint action recognition using motion history volumes”, pp. 249–257, 2006.

[67] J. Pang, Q. Huang, S. Yan, S. Jiang, and L. Qin: “Transferring Boosted Detectors Towards Viewpoint and Scene Adaptiveness.”, *IEEE Transactions on Image Processing*, Vol. 20, No. 5, pp. 1388–1400, 2011.

[68] 土屋成光, 弓場竜, 山内悠嗣, 山下隆義, 藤吉弘亘: “共変量シフトに基づく Transfer Forest”, 電子情報通信学会 パターン認識・メディア研究会, 第 114 巻, pp. 31–36, 2012.

[69] F. Shen, H. Yu, K. Sakurai, and O. Hasegawa: “An Incremental Online Semi-supervised Active Learning Algorithm Based on Self-organizing Incremental Neural Network”, *Neural Comput. Appl.*, Vol. 20, No. 7, pp. 1061–1074, October 2011.

[70] L. Breiman: “Random Forests”, *Mach. Learn.*, Vol. 45, No. 1, pp. 5–32, October 2001.

[71] J. Shotton, M. Johnson, and R. Cipolla: “Semantic texton forests for image categorization and segmentation”, in proceedings of *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8. IEEE, June 2008.

[72] P. C. Library. “http://pointclouds.org/”.

[73] Y. Linde, A. Buzo, , and R. Gray: “An algorithm for vector quantization design”, *IBBE TRANSACTIONS ON COMMUNICATION*, Vol. 28, No. 1, pp. 84–94, 1980.

[74] 天野敏之, 日浦慎作, 山口証, 井口征士: “固有空間照合法に基づく距離画像からの物体の姿勢検出”, 電子情報通信学会論文誌, Vol. J80-D-2, No. 5, pp. 1136–1143, 1997.

研究業績一覧

学術論文

[1] 弓場 竜, 三好 雅則, 村井 泰裕, 藤吉 弘亘, “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”, 電子情報通信学会論文誌, Vol. J94-D, No. 7, pp. 1101-1112, 2011.

[2] 弓場 竜, 村井 泰裕, 藤吉 弘亘, “時空間特徴の統計的学習を用いた異常行動検知”, 電子情報通信学会論文誌, Vol. J95-D, No. 6, pp. 1369-1379, 2012.

[3] R. Yumiba, H. Fujiyoshi, “A Compensation Method of Motion Features with Regression for Deficient Depth Image”, Information Processing Society of Japan Transactions on Computer Vision and Applications, Vol. 6, pp. 103-110, 2014.

国際会議発表論文 (査読あり)

[1] R. Yumiba, M. Miyoshi, H. Fujiyoshi, “Moving object detection with background model based on spatio-temporal texture”, in Proceedings of Workshop on Applications of Computer Vision, 2011.

[2] R. Yumiba, Y. Agata, H. Fujiyoshi, “A Compensation Method of Motion Features with Regression for Deficient Depth Image”, in Proceedings of International Workshop on Human Activity Understanding from 3D Data (in conjunction with CVPR), 2013.

[3] R. Yumiba, H. Fujiyoshi, “Motion Features Compensation for Monitoring Actions using Deficient Depth Images” in Proceedings of Korea-Japan Joint Workshop on Frontiers of Computer Vision, 2014.

学会口頭発表 (査読なし)

[1] 弓場 竜, 三好 雅則, 藤吉 弘亘, “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”, 情報処理学会コンピュータビジョンとイメージメディア (CVIM) 研究報告, vol.171, no.22, pp.1-8, 2010.

[2] 弓場 竜, 縣 禎輝, 藤吉 弘亘, “回帰推定による欠損距離画像の動作特徴量補正法 ”, 電子情報通信学会 パターン認識・メディア理解研究会 (PRMU) 研究報告, vol.112, no.197,pp.37-42,2012.

解説記事

[1] 弓場竜, 藤吉弘亘, “時空間特徴の統計的学習を用いた異常行動検知”, アドコム・メディア O plus E : Optics・Electronics, 35(6), pp. 604-611, 2013.