Vol.2011-CVIM-177 No.19
2011/5/19

姿勢変動を考慮した 基幹リンクモデルによる複数人物追跡

橋本潔[†] 加賀屋智之[†] 片岡裕雄[†] 里雄二[††]
田靡雅基[††] 大島京子[††] 藤田光子[††] 青木義満[†]

画像情報からの人物の姿勢推定技術は，セキュリティシステムにおける行動認識，スポーツ映像のフォーム解析など，様々な応用が期待されているコンピュータビジョンの重要課題である．本稿では，高精度な姿勢推定につなげるための前処理として人物追跡手法を提案する．ここでは，部分的な遮蔽や姿勢変化に対し，安定した追跡を実現するために，対象を頭，胴体などの基幹部位毎に分割して追跡する．さらに，各部位間に「人間的なつながり」を拘束条件に組み込むことで，効率的に探索を行う．基幹部位毎の追跡により対象の大まかな姿勢を推定することで，後の姿勢推定処理で有益な情報を取得することが可能となる．従来手法との性能比較実験を行い，提案手法の有効性を示した．

Multi-human Tracking with Main-Parts-Link-Model Considering Postural Change

Kiyoshi Hashimoto[†] Tomoyuki Kagaya[†]
Hirokatsu Kataoka[†] Yuji Sato[††] Masamoto Tanabiki[††]
Kyoko Oshima[††] Mitsuko Fujita[††] and Yoshimitsu Aoki[†]

The technique of human pose estimation from videos is very challenging problem. That enables us to recognize activities in security system and to analyze forms in sports movies. In this paper, we propose a novel method of human tracking as a preprocessing of high-accuracy pose estimation. Our system tracks human main parts such as head, torso. In addition, we constrain the position of parts with a human-like skeleton configuration to search effectively. It is able to acquire important information to limit the searching space by main-parts-tracking. To show the effectiveness of our proposed method, we perform an experiment to compare our system with the previous method.

1. はじめに

画像情報からの人物の姿勢推定技術は，セキュリティシステムにおける行動認識，スポーツ映像のフォーム解析，ジェスチャを用いたインターフェイスなど，様々な応用が期待されているコンピュータビジョンの重要課題である[1]．人物姿勢推定では，画像上で対象人物の身体各部位の位置姿勢を二次元で推定する 2D の姿勢推定と，画像上で推定した身体各部位の位置姿勢から三次元での姿勢推定を目指す 3D 姿勢推定に関する取り組みが，複数視点カメラ，もしくは単眼カメラ映像を対象として研究されている．例えば，高速に距離画像が得られるデバイスを用いるか，初期化や 3D モデルに拘束条件加えることで，リアルタイムで 3D 姿勢推定を行う手法[2, 3, 4]や，学習による識別器を用いて各部位の位置を画像上で求めることで，対象の姿勢を 2D あるいは 3D で推定する手法[5, 6]などがある．しかし，単眼カメラ映像からリアルタイムで高精度な 3D 姿勢推定をする汎用性の高いシムテムはまだ存在しない．我々は，実応用の観点から，単眼カメラ映像から人物の 3D 姿勢を推定し，動作単位を識別することで，動作単位の組み合わせとしての行動認識をリアルタイムで実現するシステムを目指している．3D での姿勢推定のため，画像上での身体各部位の位置姿勢推定結果と合わせて 3D の人体骨格モデル及び動作データベースを併用するアプローチ[7]を採用している．従来研究でもこのようなアプローチは存在するが，前述したような姿勢推定における諸問題に対応し，かつリアルタイム性を有した実用的なシステムは未だ存在しない．特に，2D 姿勢推定結果と 3D の人体骨格モデルとの対応付けの問題を解くためには，膨大な自由度をもつ探索空間中から解となる姿勢候補を絞り込む必要がある．

一般環境下におけるロバストな人物姿勢推定のためには，複雑背景下における人の切り出し，シーン中の他の人物による遮蔽，対象人物自身の人体部位による自己遮蔽などへの対応が必要となる．一般的に人物領域推定は，機械学習により生成した識別器を用いて画像上で様々な位置，スケールで探索することで行われる[7]．この処理は比較的高精度ではあるが計算コストが高く，毎フレームで検出により人物領域を推定することは現実的に有効であるとは言えない．そこで検出後は，色や形状，運動モデルなどの簡単な情報を基に対象の位置を探索し，追跡することが行われる．高精度に追跡するために，遮蔽を検知して遮蔽に対応できるような特別な処理を行う手法[8]や検出と追跡の処理を統合し安定した追跡を行う手法 [9]が提案されている．従来では，人物全体を 1 つの領域として捉えて追跡する手法が多いが，それでは部分的な遮

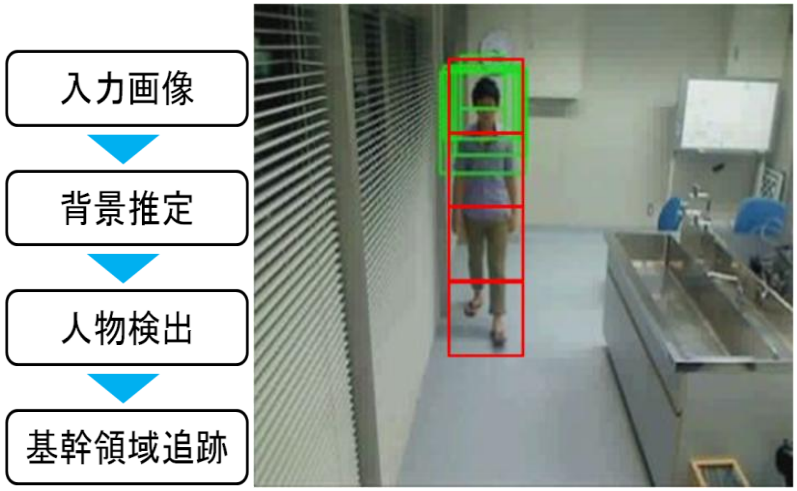
[†] 慶應義塾大学　理工学研究科
Keio University Graduate School of Science and Technology
[††] パナソニック株式会社　東京 R&D センター
Panasonic Corporation Tokyo R&D Center

Vol.2011-CVIM-177 No.19
2011/5/19

蔽が発生したときに全体の尤度が低下し，追跡に失敗してしまう．本稿では，対象を基幹領域毎に分割して追跡することで，遮蔽や姿勢変化に対して頑健な追跡手法を提案する．

2. 提案手法

我々は，実環境下におけるリアルタイムでの人物 3D 姿勢推定や，それを実現するための前処理として，単眼カメラ映像中からの高精度な人物検出手法，及び姿勢推定において強力な拘束条件となり得る人体基幹部位（頭部，胴体，腰部，脚部）の追跡手法を提案してきた[10, 11]．[10]では，部分的な遮蔽や姿勢変化に対し，基幹部位毎に追跡することが有効であることを示した．しかし，色などの限られた情報から基幹部位という特徴の少ない領域を追跡することは困難である．また従来では，各基幹部位を独立に追跡していたため，特に頭部は個人間の差異が少ないことから，複数の人物が存在するシーンでは追跡に失敗しやすいという問題があった．そこで本稿では基幹部位追跡に，部位間の「人間的なつながり」を組み込んだモデル（以下では基幹リンクモデルと呼ぶ）を導入する．これにより，基幹部位毎に独立して尤度評価をしつつ，対象を人物全体として捉えながら追跡することが可能となる．以下では，提案手法の概要について説明する．



(a) (b)

図 1 処理手順と追跡の初期化

図 1(a)に大まかな処理の流れを示す．まず初めに，背景推定と人物検出により基幹部位追跡に必要な情報を取得する．背景推定では，画像中の移動体領域を抽出する．移動体領域のみに検出処理を行うことで，探索範囲を大幅に削減することが可能となり，誤検出を低減しつつ，計算コストを大幅に削減できる．人物検出は機械学習により生成した識別器を用いて行う．検出対象は，人物の頭部から肩にかけてのΩ形状のエッジパターンとしている．これ部位を選択したのは，対象の向きや姿勢変動による形状変化や個人間の差異が少ないという理由からである．機械学習や特徴量の詳細については[10]を参照されたい．次に，検出結果を基に追跡の初期化を行う．図 1(b)では緑色の矩形が検出結果を表しており，それを統合することで一番上の赤色の矩形を得る．これを頭部のモデル画像とし，さらにその矩形を下方向にスライドさせることにより，胴体，腰部，脚部のモデル画像を取得する．この情報を使って，パーティクルフィルタにより対象を基幹部位毎に追跡していく．次章では，このときに用いる基幹リンクモデルについて説明し，尤度評価などの追跡の詳細については 4 章で説明する．

3. 基幹リンクモデル

基幹部位追跡において，各部位が画像上でどの位置にあるかを推定するために各部位の位置を独立して探索すると，追跡する部位数に対して指数関数的に探索空間が増大する．しかし，この空間内には実際の人間には骨格的に取り得ない姿勢が多く存在している．よって，部位間の拘束条件に，探索空間の無駄な範囲を絞り込むような制約を組み込むことができれば，妥当な姿勢空間のみを探索することで，計算コストを削減し，大きな精度向上が見込まれる．従来では，単純に位置が離れすぎないような距離の拘束条件を用いてつながりを表現していたが，これでは，本質的でなく「人間的なつながり」を拘束条件に組み込めていない．以下では，本稿で提案する姿勢変動を考慮した部位間のつながりを表現した新しい基幹リンクモデルについて解説する．

3.1 姿勢変動を考慮した低次元モデル

基幹リンクモデルには，部位間の拘束条件に「人間的なつながり」をうまく組み込まなければならない．本研究では，事前に取得した大量の姿勢情報が，探索空間内でどのような分布をしているかを解析し，さらに，その分布を最もよく表す低次元空間に姿勢情報を射影する．この射影を部位間の拘束条件として基幹リンクモデルとすれば，画像上での姿勢変動を効率よく表現した低次元モデルとなり，探索する空間の次元を効率よく削減できる．これにより，探索の精度と計算コストを飛躍的に向上させることが可能となる．

3.2 主成分分析によるモデル作成

主成分分析では，統計的に設定した総合的指標を数学的に合成することで，データ

Vol.2011-CVIM-177 No.19
2011/5/19

に含まれる次元間の関係を把握することが可能である．例えば，人間が様々な姿勢を取った時，各部位は骨格的な拘束条件に従った相関を持った変位をする．図 2 は，直立状態から少し前傾姿勢になったときの姿勢変化を表している．このときの画像上での各部位の変位を見ると，頭部を前に出すと同時にバランスを取るために腰部を少し後ろに下げるということが分かる．この前傾姿勢に対する変位の相関は，頭部位置の水平成分と腰部位置の水平成分の線形結合により合成された新たな次元を考えることで，1 次元で前傾姿勢の度合いを表現することが可能となる．このように，姿勢変動を含んでいる大量の姿勢データに対し主成分分析を行うことで，各部位の位置を独立に扱うのではなく，前傾姿勢など姿勢変動を表現するのに都合のいい新たな指標を導入することができる．主成分分析により射影された空間は，姿勢変動を表すのに効率的な次元で表現されている．この各次元には，元の空間に対して分析に用いた姿勢データをどれくらい表現可能かを表す重みがついている．各次元の重みを見れば，姿勢変動を効率よく表現できる次元のみを選択することで，次元削減が容易にできる．以下では，実際に基幹リンクモデルを作成する流れを説明する．

図 2　前傾姿勢に対する各部位の位置変化

まず，様々な姿勢変化を含んだ動画に対して，各部位の座標情報を手動で抽出し，頭部が原点となるように位置を正規化する．こうして得られる 6 次元情報は， 4 つの基幹部位の相対的な位置関係を表す姿勢データとなる．この 6 次元姿勢データを大量に用意することで、その分布の特性を抽出する．本研究では，図 3 のような異なる人物の映った 3 種類の動画から 720 組のデータを取得した．これに対し主成分分析を行った結果を表 1 に示す．表 1 を見ると主成分の第 2 次元までで，元々の 6 次元姿勢データの 96%以上を表現できている．そこで本システムでは，この 2 次元のパラメータによって表現された姿勢空間を基幹リンクモデルとした．このように作成された基幹リンクモデルでは，パラメータの値を変えることで人間らしい姿勢を保ったまま様々な姿勢が表現可能な低次元モデルとなっている．

表 1　主成分に対する累積寄与率

主成分	累積寄与率[%]
1	54.9
2	96.1
3	98.4
4	99.5
5	99.8

図 3　画像上から取得した各部位の位置情報

4. 基幹部位追跡

ここでは，前章で作成した基幹リンクモデルを導入したパーティクルフィルタによる基幹部位追跡について説明する．基幹部位追跡では，人物を頭部，胴体，腰部，脚部の 4 領域に分割し，それぞれを追跡することで部分的な遮蔽や姿勢変化にも柔軟に対応することが可能である．探索にはパーティクルフィルタを用いており，これは，推定したい状態空間内に多数のパーティクルを散布し，その後，各パーティクルにおいて尤度計算を行い，その重み付き平均で状態を推定する手法である．このパーティクル散布，重み計算，状態推定を繰り返すことで効率的に状態空間を探索することが可能となる．以下では，パーティクルフィルタの設定について詳細に説明していく．

4.1 推定する状態とシステムモデル

本システムで推定する状態空間は，姿勢変化を表現する基幹リンクモデルの 2 次元と画像上の位置と速度を表現する 4 次元を合わせた 6 次元空間とした．この状態空間内での最尤値を推定することで，画像上では対象の位置と大まかな姿勢を同時に最適化したことに相当する．

パーティクルフィルタでは，以下に定義したシステムモデルに従ってパーティクルを散布する．歩行中の人物のような移動物体は，簡単な運動モデルを当てはめることが可能であるが，姿勢変化のように複雑な高次元空間を移動するものはモデルの設定が難しいため，システムモデルは以下のように定義した．画像上の人物は等速に移動すると仮定し，位置空間に対しては等速直線運動モデルを適用した．姿勢空間に対しては，前フレームの推定位置の周辺をランダムサンプリングするランダムウォークとなっている．このシステムモデルを用いることで，状態空間内で複雑に移動する対象に対し，効率的にパーティクルを生成できる．状態推定では，最尤値を全パーティクルの尤度による重み付き平均値とすることで行われる．

$$\mathbf{x}_t = (\,PC_1\ PC_2\ x_t\ y_t\ \dot{x}_t\ \dot{y}_t\,)^T \tag{1}$$

$$\mathbf{F} = \begin{bmatrix} 1&0&0&0&0&0 \\ 0&1&0&0&0&0 \\ 0&0&1&0&1&0 \\ 0&0&0&1&0&1 \\ 0&0&0&0&1&0 \\ 0&0&0&0&0&1 \end{bmatrix} \tag{2}$$

$$\mathbf{x}_t = \mathbf{F}\mathbf{x}_{t-1} + \mathbf{v}_t^{(i)} \tag{3}$$

$\mathbf{x}_t$：状態ベクトル
PC_i：基幹リンクモデルに用いた主成分分析の第 i 主成分
x_t, y_t：対象の位置
$\dot{x}_t, \dot{y}_t$：対象の速度
$\mathbf{v}_t^{(i)}$：システムノイズ

4.2 尤度評価

この処理では，各パーティクルの重みを決定する．この重み付き平均で状態を推定するため，追跡精度を決定する重要な処理である．各パーティクルは，画像上で 4 つの基幹部位がどの位置にあるかを表現する 6 次元の情報を持っている．本システムでは，全基幹部位に対しそれぞれ尤度計算を行い，その合計をパーティクル全体の尤度とする．尤度計算は，3 つのステップにより行われる．まず，各基幹部位位置に対し，その周辺画像を取得する．その後，頭部以外の部位は，初期化時に取得したモデル画像と色ヒストグラムを比較することで，その類似度を尤度とする．ヒストグラム比較には 4 式の Bhattacharyya 距離を用いた．色ヒストグラムのビン数は 180 とし，2 つのヒストグラムサイズは画像サイズによらず 10000 に正規化してある．頭部の尤度は検出処理で用いた識別器によって評価する．頭部画像からエッジベースの特徴量を抽出し，特徴ベクトルを識別器に入力することで尤度を計算する．学習には AdaBoost を用いたので，その出力は 0~1 となり，その値をそのまま尤度とした．弱識別の数は 500 としている．識別器を用いて評価することで，頭部は他の部位に比べ高精度な出力を得ることができる．また，追跡対象のスケールが変化するような場合，色ヒストグラム類似度では評価が難しい．しかし，識別器を用いることで，スケールに対しても安定した評価が可能となる．最後に全部位の尤度を合計することで，そのパーティクルの重みを計算する．

$$\boldsymbol{L}_{\mathrm{color}} = \sum_{\mathrm{bin}=1}^{180} \sqrt{p(\mathrm{bin}) \cdot q(\mathrm{bin})} \tag{4}$$

$$\boldsymbol{L}_{\mathrm{head}} = \sum_{t=1}^{500} \alpha_t h_t(x) \tag{5}$$

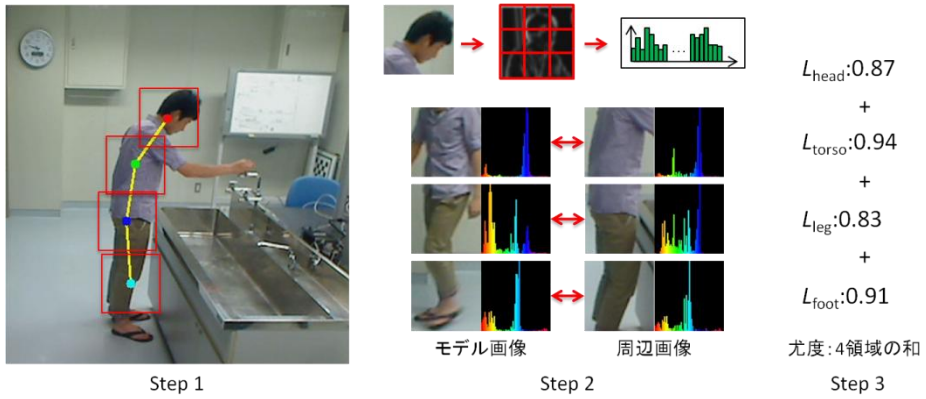


図 4　尤度計算の流れ

5. 追跡精度比較実験

5.1 実験概要

提案手法の有効性を示すために，遮蔽や姿勢変化が発生している動画に対して，その前後で追跡に成功したかを判定する実験を行った．成功の判断基準としては，人物領域を正しく推定できているか，基幹部位の位置があっているかを目視で確認した．比較のための従来手法は，[10]を用いた．これは隣接する部位間の拘束条件に，距離に反比例して尤度が低下するような単純なモデルを適用した手法であり，尤度計算は提案手法と同じものを用いた．本実験に用いた動画は，2 つのシーンで撮影した．1 つ目は図 5 のように，比較的背景が単調で照明変化も少ないシーンとなっている．2 つ目は図 6 のように，照明変化が大きく複雑な背景下のシーンとなっており，1 つ目に比べ難易度の高い動画である．結果を以下に示す．

5.2 実験結果

処理結果を見ると，従来手法ではすれ違うときに画面手前の人の影響を受けて尤度が低下することで，スケールや位置が安定しない場面が多くみられた．また，シーン 1 では両手法とも高い成功率となっているが，シーン 2 では従来手法の成功率が大きく低下している．つまり，提案手法では，複数のシーンに対して安定して高い追跡成功率を実現していることがわかる．そもそも従来手法では，基幹部位を独立に追跡しているため，複雑なシーンでは全体的にどの部位の尤度も低くなってしまうことにより，追跡に失敗している．しかし，提案手法では基幹リンクモデルによる部位間の強力かつ柔軟な拘束条件を用いることで，高精度に追跡できている．

表 2　追跡成功率

	Scene1	Scene2	合計
従来手法	73.9% (17/23)	18.8% (3/16)	51.3% (20/39)
提案手法	91.3% (21/23)	93.8% (15/16)	92.3% (36/39)

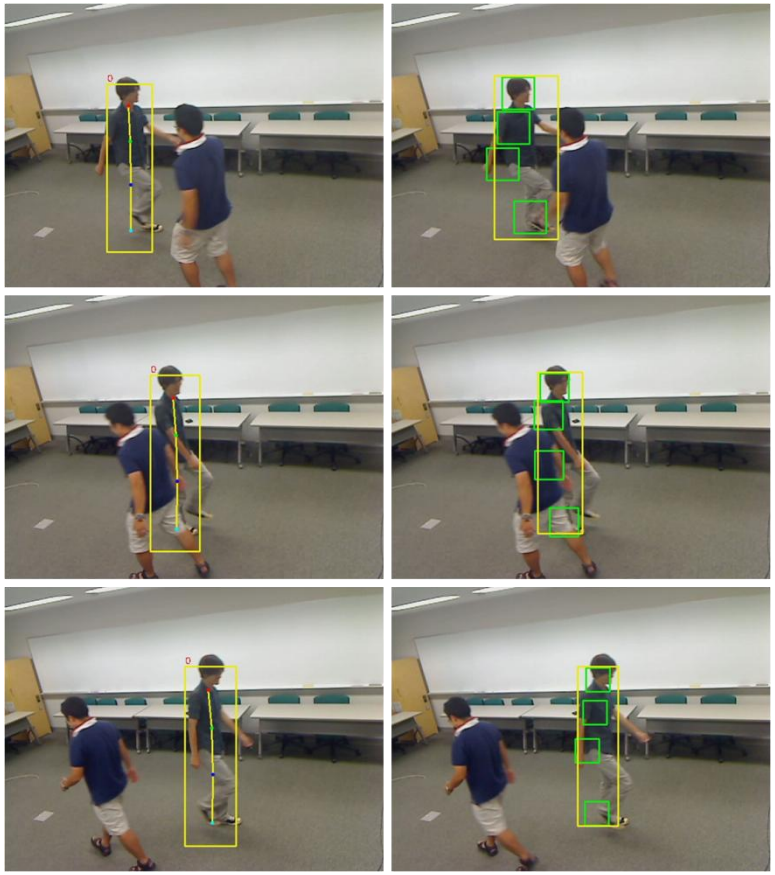


図 5　シーン 1 の追跡結果（左：提案手法，右：従来手法）

Vol.2011-CVIM-177 No.19
2011/5/19

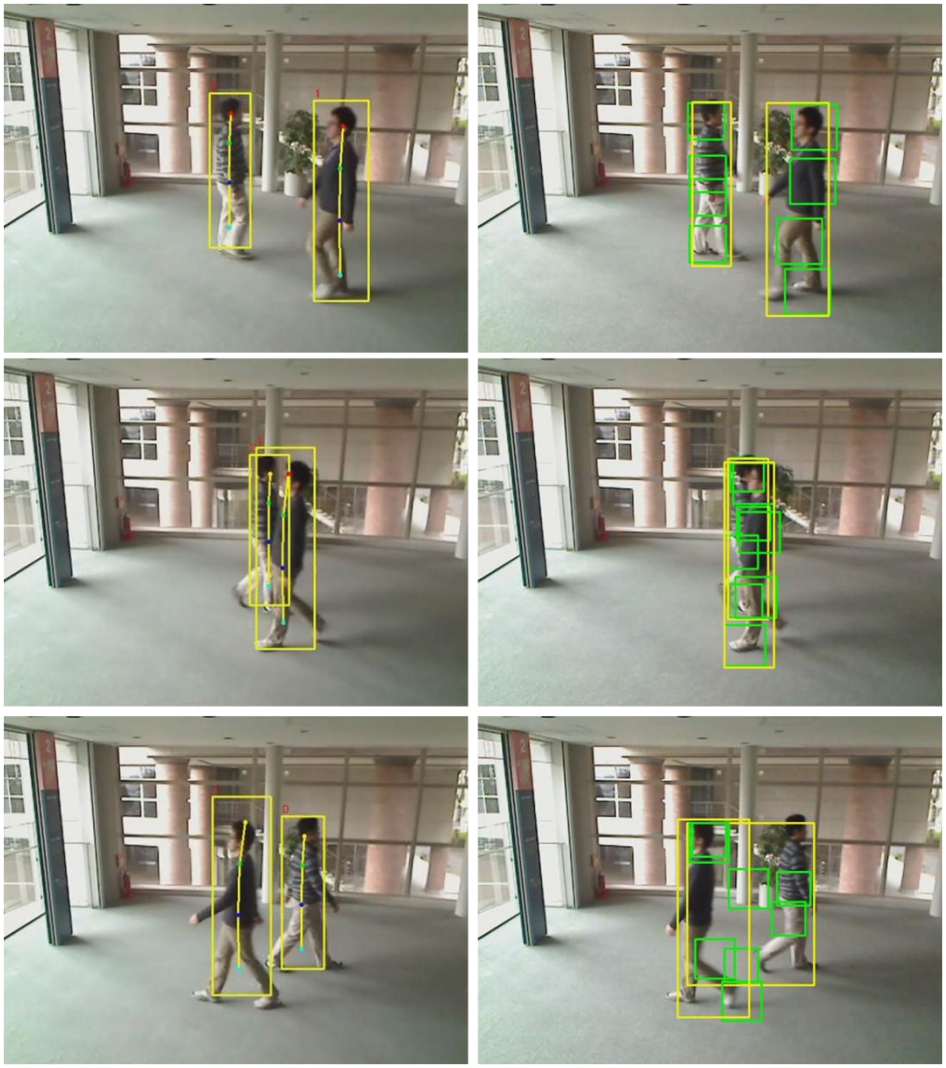


図 6　シーン 2 の追跡結果（左：提案手法，右：従来手法）

6. 考察

本研究では，効率的な基幹部位追跡を行うために新たに基幹リンクモデルを作成した．これは，各基幹部位の位置関係を表現する膨大な状態空間を主成分分析により低次元で表現したものである．図 6 は，システムモデルに従って状態空間に散布したパーティクルの基幹リンクモデルの有無による違いを表しており，見やすくするために少数のパーティクルのみを画像上の位置情報に変換して表示している．基幹リンクモデルを用いずにパーティクルを生成した場合，人間が取り得る姿勢を考慮していないため，物理的にあり得ない不自然な姿勢が含まれているのに対し，基幹リンクモデルを用いたものは，実際の姿勢に対して比較的妥当で，なおかつ，位置や姿勢に適度なばらつきを含んだパーティクルのみを効率的に生成していることが分かる．これは，従来では表現できなかった部位間の「人間的なつながり」を考慮することで，高次元の状態空間に対して，より少ない次元の空間でサンプリングすることができ，少ないパーティクル数でも十分な探索が可能となる．これにより，処理速度を落とすことなく，飛躍的に精度を向上することができたと考えられる．姿勢変化への対応として，対象全体を 1 つの矩形で追跡する場合，姿勢が変化したときに人物領域のみを柔軟に評価，推定することができなかった．本手法では姿勢変動を考慮した基幹リンクモデルを用いて基幹部位毎に追跡を行うため，様々な姿勢に対して柔軟に基幹部位周辺画像を取得することで，尤度が低下することなく安定して人物領域のみを推定することが可能となる．

図 7　基幹リンクモデルの有無によるサンプル生成の違い

基幹部位追跡の枠組みでは，部位毎に独立して追跡を行いながらも，基幹リンクモデルを用いることで，遮蔽や姿勢変化に対しても頑健に人物全体をとらえることが可能な追跡を実現した．遮蔽が発生した場合でも，部位毎に尤度評価をすることで，部分的な遮蔽に対して全体の尤度が低下することがない．さらに，尤度評価に識別器を用いることで頭部位置を安定して推定できるため，対象をまず見失わない．以上のことにより，本システムでは，遮蔽を検知してそれに対処するような特別な処理をすることなく，安定した追跡を実現した．また，基幹部位追跡では人物領域に加え，各部位の位置情報が取得できる．後の処理で高精度な姿勢推定をすることを考えると，この情報は非常に有益なものとなり，姿勢の自由度を大幅に削減することで高速かつ高精度な推定が可能となる．

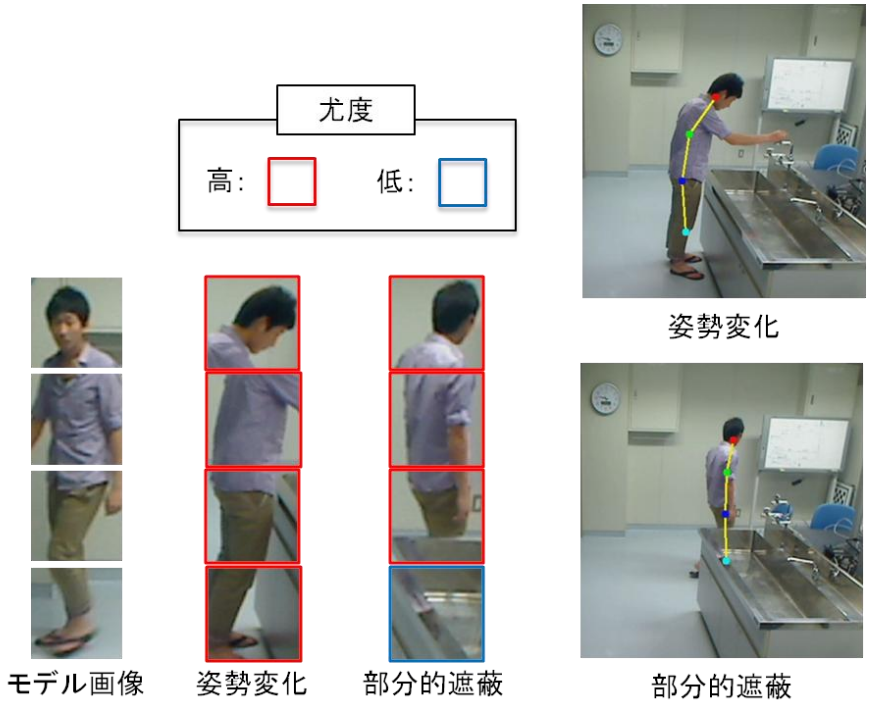


図 8　基幹部位追跡による遮蔽，姿勢変化への対応

7. おわりに

7.1 結論

本研究では，姿勢推定に適した高速かつ安定な追跡を実現することを目標としている．しかし，実環境では照明環境の変動や環境物や人物間による遮蔽，対象の姿勢変化など様々な原因がこれを難しくしている．そこで本稿では，この問題に対応するために，基幹リンクモデルによる基幹部位追跡を行うことで，遮蔽や姿勢変化に対しても高速で安定した追跡が可能な手法を提案した．提案手法では遮蔽や姿勢変化が発生する難しいシーンに対し，92%以上の追跡成功率を実現している．さらに，基幹部位追跡では人物領域と同時に各基幹部位の位置情報も推定する．この情報は姿勢推定の段階で有益なものとなり，探索する姿勢空間の大幅な削減につながることが見込まれる．

7.2 今後の課題

現在，追跡の尤度評価に識別器を用いているため，処理時間が準リアルタイム程度となっている．特徴量や学習手法を改善するなど，速化の処理を考えていきたい．また，姿勢推定において重要な情報である肩の位置や対象の向きを推定するような処理を加えていきたい．

参考文献

1) Thomas B. Moeslund, Adrian Hilton, et al.: A survey of advances in vision-based human motion capture and analysis, Computer Vision and Image Understanding 2006, pp. 90-126
2) V. Ganapathi, C. Plagemann, D. Koller, and S. Thrun: Real time motion capture using a single time-of-flight camera, Computer Vision and Pattern Recognition 2010, pp. 755, (2010)
3) Jamie Shotton, et al.: Real-Time Human Pose Recognition in Parts from Single Depth Images, Computer Vision and Pattern Recognition 2011, (2011)
4) Shian-Ru Ke, et al.: Real-Time 3D Human Pose Estimation from Monocular View with Applications to Event Detection and Video Gaming, Advanced Video and Signal-Based Surveillance 2010, (2010)
5) M. Andriluka, S. Roth and B. Schiele: Pictorial structures revisited: People detection and articulated pose estimation, Computer Vision and Pattern Recognition 2009, pp. 1014-1021, (2009)
6) Mykhaylo Andriluka, Stefan Roth, Bernt Schiele: Monocular 3D Pose Estimation and Tracking by Detection, Computer Vision and Pattern Recognition 2010, pp. 623-630, (2010)
7) Tomoki Watanabe, Satoshi Ito, Kentaro Yokoi: Co-occurrence Histograms of Oriented Gradients for Pedestrian Detection, Pacific-Rim Symposium on Image and Video Technology 2009, pp. 37-47, (2009)
8) M. S. Ryoo, and J. K. Aggarwal: Observe and Explain: A New Approach for Multiple Hypotheses Tracking of Humans and Objects, Computer Vision and Pattern Recognition 2008, pp. 1-8, (2008)
9) Michael D. Breitenstein, et al.: Robust Tracking-by-Detection using a Detector Confidence Particle Filter, International Conference on Computer Vision 2009, pp. 1515-1522, (2009)

Vol.2011-CVIM-177 No.19
2011/5/19

10) 橋本潔, 青木義満 他: 単眼カメラを用いた姿勢推定のための人物検出と基幹部位追跡, ViEW2010, pp. 257-263, (2010)
11) 加賀屋智之, 青木義満 他: 部位尤度と人体モデル照合に基づく単眼カメラ映像からの人物 3 次元姿勢推定, DIA2011, pp. 106-110, (2011)