

姿勢変動を考慮した基幹部位追跡と肩位置推定 -単眼カメラからの3次元姿勢推定にむけて-

橋本 潔[†] 加賀屋 智之[†] 片岡 裕雄[†] 里 雄二[‡]

田摩 雅基[‡] 大島 京子[‡] 藤田 光子[‡] 青木 義満[†]

[†] 慶應義塾大学 理工学研究科 〒223-8522 神奈川県横浜市港北区日吉 3-14-1

[‡] パナソニック株式会社 東京 R&D センター 〒224-8539 神奈川県横浜市都筑区佐江戸町 600 番地

E-mail: [†] khashimoto@aoki-medialab.org

あらまし 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、頭から肩にかけての外輪郭形状から肩の位置を推定する。これにより、対象の姿勢を荒く推定しながらの追跡が可能となる。

キーワード 人物追跡, 特徴量, 姿勢推定, 肩位置推定, パーティクルフィルタ, 主成分分析

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 モデルに拘束条件を加えることで、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ス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正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。また、姿勢変化に対しても柔軟に人物領域のみを捉えることができない。本稿では、対象を基幹部位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。

また、姿勢を推定する際に、上半身では、足が地面に付いているというような重力的な拘束が少ない分、取り得る姿勢の自由度はかなり高い。よって基幹部位追跡から得られる頭部や胴体の位置情報だけでは肩や腕の姿勢を推定するためのヒントとして適切とはいえない。そこで、本稿ではさらに姿勢推定処理を意識した前処理として肩位置と向きを推定する。肩位置推定では、輪郭形状の変動とその内部に含まれる肩の関節位置を持つモデルをフィッティングさせることで、肩

の位置を推定する．このように，変形可能なモデルのフィッティングにより頭や顔の形状を追跡することで，内部に持つ頭部の姿勢や顔の特徴点位置を推定する手法[10,11]は提案されているが，肩位置を推定する手法はない．

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[12, 13]．[12]では，基幹部位毎に独立して尤度評価しつつ，基幹リンクモデルにより部位間の位置関係を拘束することで人物全体を柔軟に捉え，部分的な遮蔽や姿勢変化に対しても安定した追跡を実現した．これまで，安定して人物領域を推定することを目的として基幹部位追跡手法を提案してきたが，本稿ではさらに，姿勢推定の強力な拘束条件となり得る肩位置と向きを推定する．肩位置推定では，複数の異なる人物画像から生成した輪郭形状の変動とその内部に含まれる肩の関節位置を持つ **Active Shoulder Model** を用いて，頭部のエッジやシルエットとの形状の一致度を評価値とすることで，最もフィットする位置とモデルを探索し，そのモデルが持っている肩の位置を推定結果とする．また向き推定では，モデルが持っている頭部と肩を結ぶ三角形の歪み度合いから推定する．各処理の詳細については後章で説明する．まず，3章で基幹リンクモデルによる基幹部位追跡について簡単に説明した後に，4章で肩位置推定，5章で向き推定の説明をする．

3. 基幹リンクモデルによる基幹部位追跡

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，映像中の移動体領域を抽出する．移動体領域のみを検出対象とできるため，探索範囲の大幅な削減が可能となり，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．この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[12]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．

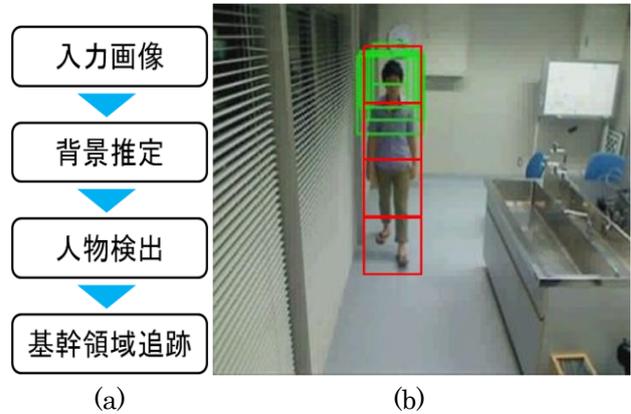


図 1 処理手順と追跡の初期化

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．そこで，部位間の拘束条件に「人間的なつながり」をうまく組み込んだ基幹リンクモデルを用いる．図 2 のように人間は様々な姿勢を取った時に，基幹部位位置は骨格や重心のバランスなどの拘束条件に従った変位をする．この関係を，様々な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，統計的に解析する．これにより，前傾姿勢など姿勢変動の表現に都合のいい新たな次元を組み込んだ基幹リンクモデルを生成した．



図 2 前傾姿勢に対する各部位の位置変化

姿勢を表す基幹リンクモデルと画像上の位置(2次元)を合わせた空間を状態空間としてパーティクルフィルタにより基幹部位毎に追跡する。パーティクルフィルタとは、推定したい状態空間内をシステムモデルに従って生成した多数のパーティクルによってサンプリングし、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。状態空間のサンプリング、尤度計算、状態推定の処理を繰り返すことで効率的に状態空間の探索、状態の推定をすることが可能となる。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のは運動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。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。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。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尤度を算出する。状態推定では、状態空間内の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して求めることで、画像上での基幹部位位置を推定する。

基幹リンクモデルを用いた基幹部位追跡では、人物を頭部、胴体、腰部、脚部の4領域に分割し、それぞれを独立に尤度評価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。また、部位間に「人間的なつながり」を組み込んだ拘束条件を用いて探索することで、効率的に高精度な探索が可能となっている。[12]では、複雑背景、複数人物、遮蔽、姿勢変化を含むシーンにおいて、93.8%の追跡精度を実現した。

4. Active Shoulder Model による肩位置推定

肩位置情報は姿勢推定において非常に有益な情報となる。ここでは、Active Shoulder Model を用いた肩位置推定手法を提案する。図3を見ると人体頭部の外輪郭形状と肩の位置は相関がありそうに見える。例えば、正面を向いていても横を向いていても肩の曲線の中心近くに肩の関節が存在する。そこで、本稿ではこのことを利用して肩位置を推定する。まず、人間らしい形状を保った変形が可能な頭部外輪郭形状を表現する Active Shoulder Model を事前に作成する。このモデルには、内部に形状変形に対応して移動する肩位置情報を持たせておく。次に、画像上で最もよくフィットするモデルの形状と位置を求めることで、同時に肩位置を推定する。

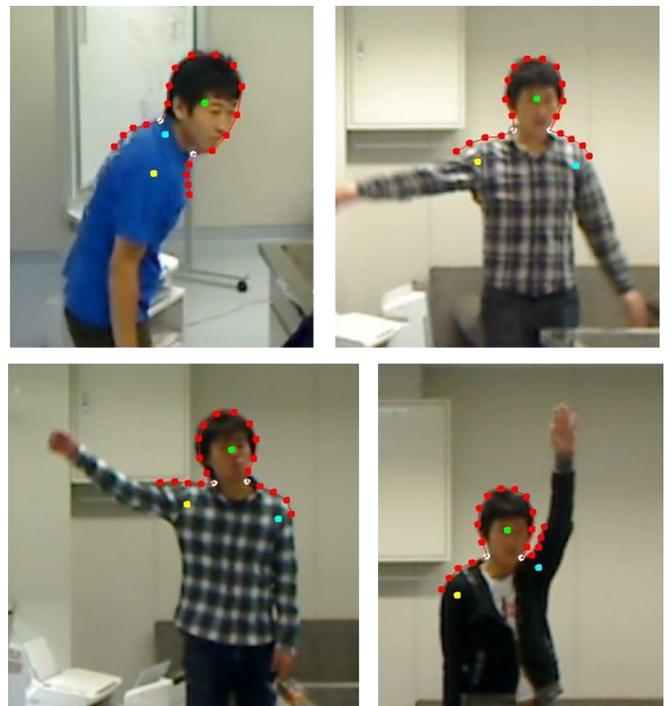


図3 頭部の外輪郭形状データ

4.1 Active Shoulder Model の生成

Active Shoulder Model (以下、ASM と省略する) とは、輪郭形状の変動とその内部に含まれる肩の関節位置を持つ低次元モデルであり、以下の手順で生成する。まず、図3のような複数の異なる人物頭部画像に対して、顔位置を中心として頭部から肩にかけての外輪郭形状と両肩の関節位置を合計 22 個の点群で表現する。22 点の x, y 座標から得られる 44 次元の頭部外輪郭形状データを主成分分析することで、低次元で向きや姿勢による頭部外輪郭形状の変化を表現でき、それに対応して肩位置が移動するモデルとなる。本稿では 154 個の頭部外輪郭形状データに対し、主成分分

析を行うことで、3次元に圧縮した。この主成分分析の部分空間をパラメータとするモデルをASMとした。

4.2 ASM フィッティング

画像上で最適なASMを探索することで、肩位置を推定する。探索には、パーティクルフィルタを用いた。手順は次のようになる。まず、ASM(3次元)と画像上の位置(2次元)の計5次元を状態空間として、多数のParticleでサンプリングする。システムモデルには、ランダムウォークを用いており、尤度計算は式(1)で行った。PFで毎フレームの頭部外輪郭形状を追跡フィッティングすることで、ASMの持っている肩位置を推定結果とする。

$$L(\Omega) = \frac{1}{1 + C_1 \cdot S_{\text{edge}}(\Omega)} + \frac{1}{1 + C_2 \cdot S_{\text{sil}}(\Omega)} \quad (1)$$

$$S_{\text{edge}}(\Omega) = \int_{s(x,y) \in \Omega} |d_i(x,y) - d_{\text{ASM}}(x,y)| ds \quad (2)$$

$$S_{\text{sil}}(\Omega) = \int_{s(x,y) \in \Omega} D(x,y) ds \quad (3)$$

$L(\Omega)$: 集合 Ω によって表させるASMの尤度

Ω : ASM上にある点の集合

$s(x,y)$: ASM上のある点

$d_i(x,y)$: 入力画像Iの点(x,y)における勾配方向

$d_{\text{ASM}}(x,y)$: ASM上の点(x,y)における法線方向

$D(x,y)$: 点(x,y)における距離変換画像の画素値

C_1, C_2 : 正規化のための係数

ここで、 S_{edge} と S_{sil} はそれぞれ頭部の輝度勾配と背景差分画像のシルエットから求めた評価値で以下のようにして求める。また、 C_1 と C_2 はエッジによる評価とシルエットによる評価の重みが1対1になるよう正規化するための係数である。エッジによる評価では、評価したいASM上の各ピクセルにおいて、ASMの法線方向と画像の輝度勾配方向の差を合計したものを評価値とする。これにより、ASMの形状が頭部の輪郭の形状に一致している程、評価値は小さくなる。シルエットによる評価では、まず、背景差分画像において、そのまま距離変換したものと背景と前景を反転させてから距離変換したものを組み合わせることで得られる図4のような画像を生成する。この画像は、背景差分画像における前景領域と背景領域の境界線からの距離を画素値としたものになっている。よって、評価したいASM上の画素値を積分することで得られる評価値は、そのASMがシルエットに一致している程、小さ

い値になる。

エッジとシルエットそれぞれから求めた評価値の和をASMの尤度とすることで、背景エッジの影響を受けることなく、安定した頭部外輪郭形状の尤度評価が可能となる。フィッティング結果および肩位置推定結果を6章に示す。

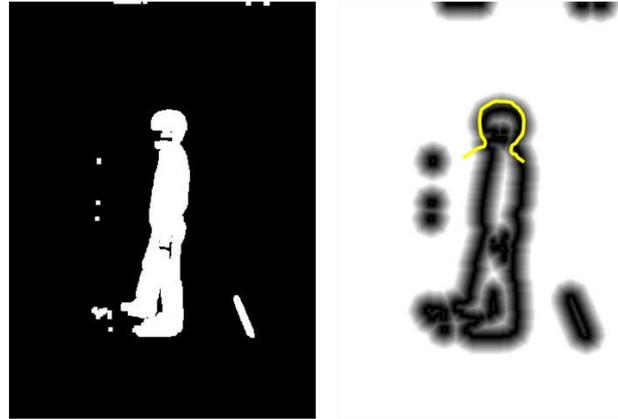


図4 背景差分画像とその距離変換

5. 向き推定

ここでは、ASMが持っている頭、両肩の3点を結ぶ三角形を利用して向きを推定する。この三角形の頂角は横を向くに従って小さくなり、真横を向いたときに 0° になる。よって、正面時の頂角の大きさを基に、現在の頂角の大きさとの比を求めることで向きが推定できる。表1は、97人の人体寸法データ[14]から顔の高さと肩幅の関係についてまとめたものである。

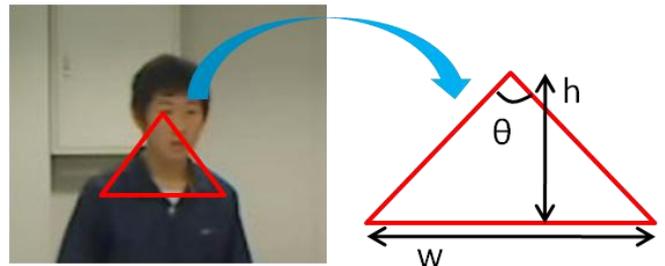


図5 頭部と両肩を結ぶ三角形

表1 肩幅と顔の高さの関係

	肩幅[mm]	顔の高さ[mm]	頂角[°]
最大	347	253.5	68.8
最小	272	355	55.3
平均	344.8	298.3	60.1

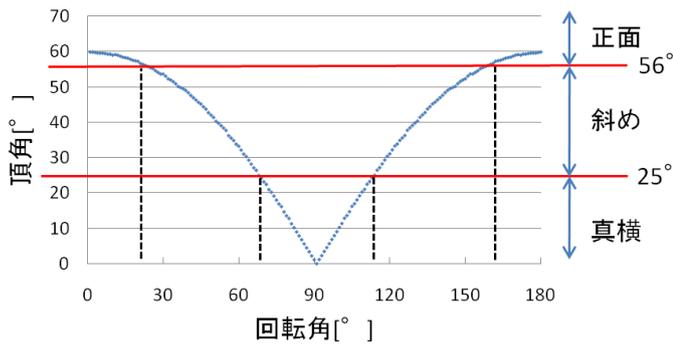


図 6 向きの量子化

表 1 より頂角の最大, 最小および平均の値を見ると, 頭部と両肩を結ぶ三角形の大きさには個人差が存在する. そこで, 推定する向きを正面(0°), 斜め(45°), 真横(90°)の 3 方向に量子化することでこの誤差を吸収する. 図 6 は, 頂角 60° の三角形をカメラに対して水平方向に回転させた時の頂角と回転角の大きさの関係を示している. 図 6 のように量子化することで頂角の大きさに対する向きを推定する. また, この方法では正面と後ろ向きの判定はできないが, これは対象の移動方向などから推定可能であるため問題ではない. 推定結果を図 7 に示す. 図のカラーチャートは人物を真上から見たときに向いている方向による推定結果の表示色の分布を示しており, 対象の向きに応じて, 正しく向きが推定できていることが分かる.

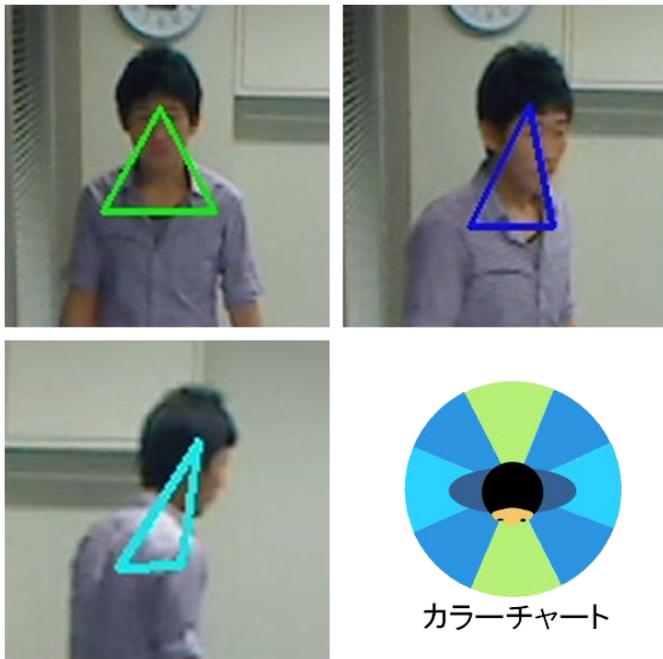


図 7 向き推定結果

6. 実験

肩位置推定の精度を検証するために実験を行った. 向きや姿勢の変化が発生している様々な動画に対し, ASM フィッティング後のモデルの肩位置が, 手動で抽出した正解の肩位置から一定の範囲内にあるかどうかを毎フレームで判定した. 本実験では映っている頭部の大きさから 6 ピクセル以内を成功とした. 評価に用いた動画には, 腕を動かすといった姿勢の変化 (動画 1,2) と向きや傾きによる見えの変化 (動画 3) が含まれている. 結果を以下に示す.

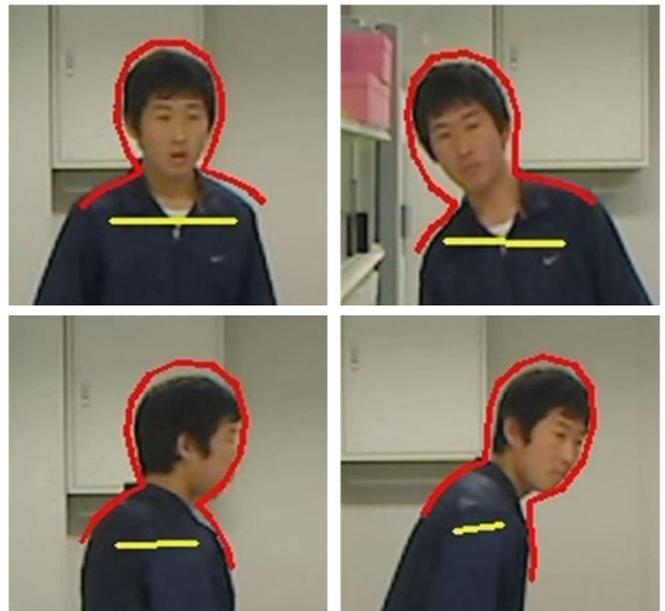


図 8 ASM による肩位置推定結果

表 2 肩位置推定精度

動画	成功[frame]	失敗[frame]	精度[%]
1	97	17	85.1
2	185	67	73.4
3	501	42	92.2

図 8 より様々な姿勢において頭部外輪郭形状を ASM のフィッティングにより安定して推定できている. また, 表 2 より肩位置の推定精度も高いことが分かる.

7. 考察

ここでは ASM による肩位置推定について考察する. 本手法では, 事前に多数の頭部外輪郭形状データから生成した ASM を用いて PF で探索することで, 効率的に頭部外輪郭形状を推定できている. フィッティングのための尤度計算には, エッジとシルエットから求めた 2 つの評価値を用いている. エッジによる評価では, 画像の勾配方向を見て ASM との形状の一致度を評価

しているが、これだけでは、背景に複雑なエッジがあるときにうまく評価できない。そこで、頭部外輪郭周辺のエッジのみを評価するために、背景差分画像から得たシルエットによる評価を加えることで背景の影響を受けない安定した尤度計算が可能となる。これにより、様々な姿勢による頭部外輪郭形状を頑健にフィッティングできている。前章の実験により ASM がうまくフィッティングできていれば、その形状から安定して肩位置を推定可能であること示した。しかし、対象が手を大きく上にあげたときでは、ASM の持っている形状空間内では表現できない程に頭部外輪郭形状が変形してしまい、フィッティングができないという問題がある。これに対しては、尤度の評価方法を改善するなどして対応していきたい。

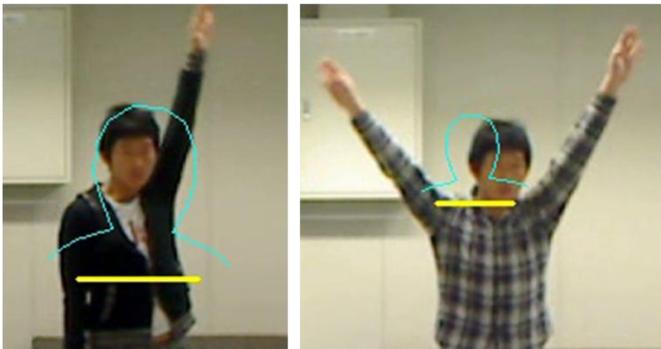


図5 フィッティングの失敗例

8. まとめ

本稿では、姿勢推定のための前処理として、人物領域だけでなく、対象の大まかな姿勢（頭部、胴体、肩など基幹部位置情報）まで同時に推定する手法を提案した。これにより姿勢推定の探索自由度を大幅に削減することが可能となる。特に、ここで推定した肩位置は、上半身や腕の姿勢を推定する際に非常に有益な情報となる。また、現在のシステムでは基幹部追跡と肩位置推定を独立に行っているが、互いの情報を使って補完し合えるよう2つの処理を統合することでシステム全体の精度を向上させていきたい。

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)

[10] Michael Isard, and Andrew Blake: CONDENSATION – Conditional Density Propagation for Visual Tracking, *International Journal of Computer Vision*, vol. 29, No. 1, pp. 5-29, (1998)

[11] F. Dornaika and J. Ahlberg: Efficient Active Appearance Model for Real-Time Head and Facial Feature Tracking, *Analysis and Modeling of Faces and Gestures* 2003, pp. 173-180, (2003)

[12] 橋本潔, 青木義満 他: 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, *CVIM 5月研究会*, Vol.2011, CVIM177, No.19, (2011)

[13] 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)

[14] AIST/HQL 人体寸法・形状データベース 2003 <http://riodb.ibase.aist.go.jp/dhbodydb/fbodyDB/index.html>

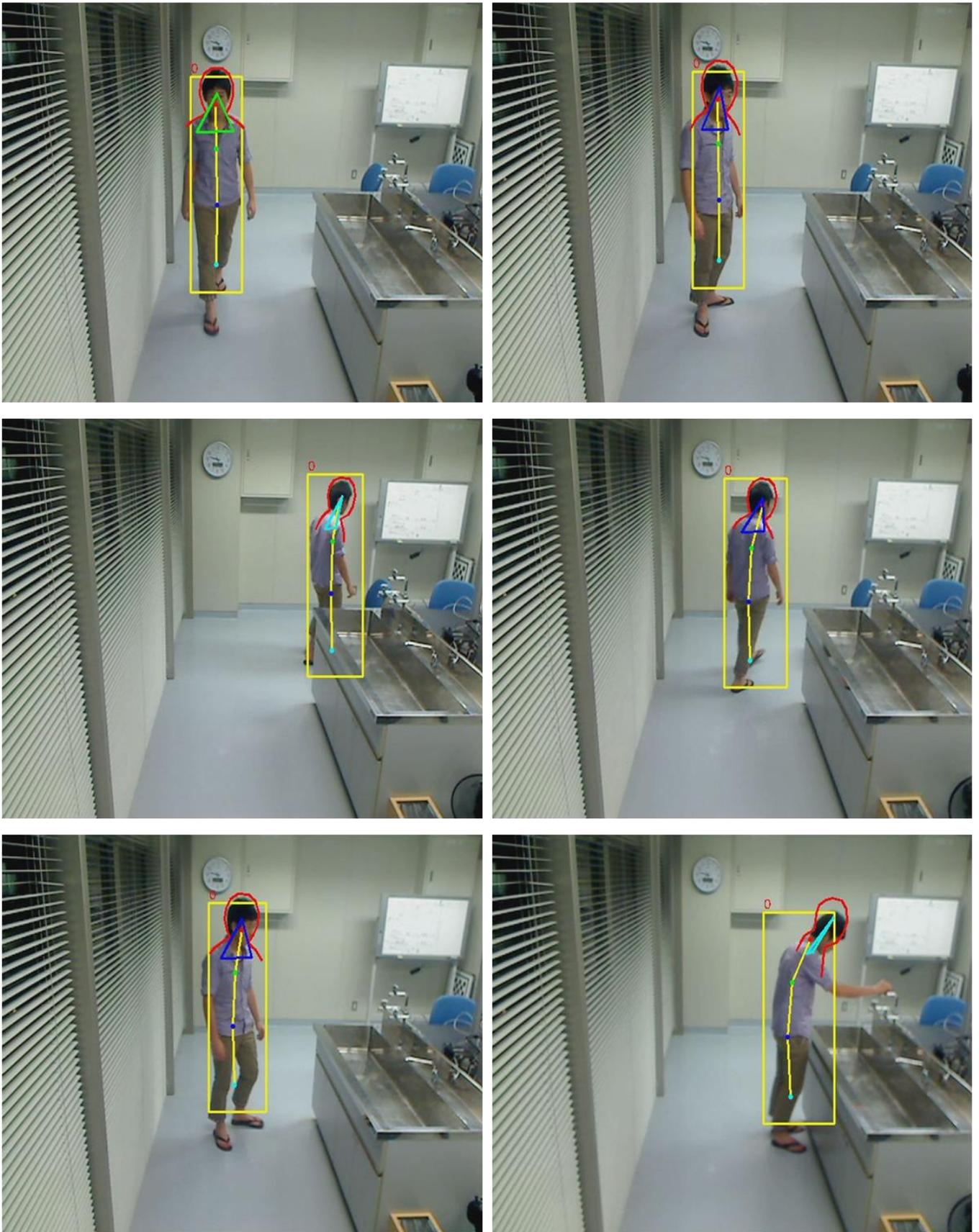


図9 基幹部位追跡，肩位置推定，向き推定の処理結果