

関節物体のモデルフィッティング

島田 伸敬[†] 有田 大作^{††} 玉木 徹^{†††}

本稿では人体や手などに代表される多関節物体の三次元姿勢を画像から推定するモデルフィッティングの技術についてサーベイする。画像によるモデルフィッティングの枠組みを、1) 推定に利用される画像特徴、2) 画像と照合するモデルの表現と照合のパラメータ空間、3) 照合時の評価関数と最適解の探索手法、にわけて多関節物体の三次元姿勢推定に特徴的な要素を上記三つの観点から比較整理することを試みる。

Model Fitting of Articulated Objects

NOBUTAKA SHIMADA,[†] DAISAKU ARITA^{††} and TORU TAMAKI^{†††}

In this paper, we present a survey report for the model fitting method to estimate 3-D posture of articulated objects such as human body and hand. We decompose the model fitting framework into the following three elements: 1) image feature, 2) model description and parameter space for model-image matching and 3) matching function and its optimization. From the viewpoint of these three issues, we try to compare the various methods of model fitting to each other and summarize them.

1. はじめに

画像から物体の姿勢を推定する問題はコンピュータビジョンにおける基本問題として古くから取り組まれている。とくに近年は人物認証や保安用途、HCIといった分野において、人間をみるビジョン技術のニーズが高まっている。本報告では人間のような多関節物体の三次元姿勢を画像から推定する手法についてサーベイする。一般にジェスチャや動作を識別する場合、必ずしも三次元姿勢の計測・推定が必要とは限らないが、ここではモーションキャプチャから動作認識までの広い応用分野において要素技術となりうる、画像による姿勢推定の技術についてターゲットとした。

本報告をまとめるにあたり、画像による三次元姿勢推定の枠組みを以下のように一般化してとらえることにした。

- (1) 推定に利用される画像特徴
 - (2) 画像と照合するモデルの表現と照合のパラメータ空間
 - (3) 照合時の評価関数と最適解の探索手法
- 画像からは、対象の姿勢を決める手がかりとしての特

徴が抽出される。一方、当てはめるべきモデルには、対象物体をどのように記述するかによってさまざまな表現が考えられるが、モデルと画像特徴間の照合を行うには、本質的になんらかの共通空間に両者を写像して比較するプロセスが内在するといえる。つまり特徴の対応付けや、最小二乗法における残差や確率的照合における尤度、事後確率の定義といったものである。これらによってモデルと画像特徴間の照合度が与えられれば、姿勢推定問題はいわゆる一般のモデル当てはめ問題としてとらえることができる。

上記三つの要素どうしはそれぞれ密接にかかわっており、どのような画像特徴と照合するかによってモデルの表現がきまる（またはその逆）、さらにそれらが照合時の評価関数の性質に影響するため、それらに応じて最適解の探索手法が選択される、といった具合である。

本報告では主に多関節物体の三次元姿勢推定に特徴的な要素を上記三つの観点から比較整理することを試みる。2節では、当該分野における従来のサーベイ論文を概観する。3、4、5節ではそれぞれ、画像特徴、モデルの表現とパラメータ空間、最適解の探索手法の各観点から代表的な手法について述べる。6節では当該研究分野の研究動向を踏まえて、とくに最近の研究動向の展望について述べる。最後に7節で本報告のまとめを述べる。

[†] 立命館大学
Ritsumeikan University
^{††} 九州大学
Kyushu University
^{†††} 広島大学
Hiroshima University

2. 過去のサーベイ

人体や手の動作など多関節物体の運動解析についてのサーベイは、これまでに多数行われている(169),170),97),119),4),40),117),6),3),18),5)。

おそらくその先駆けである Aggarwal ら^{5),6)}のサーベイは、Huang らの非剛体物体運動の分類^{42),50),67)}に基づいて、物体の運動を rigid, non-rigid (restricted (articulated, quasi-rigid, isometric, homothetic, conformal), general (elastic, fluid)) に分類した。非剛体の物体や運動全般を扱っている Huang らとは異なり、Aggarwal らは人体動作のような articulated と心臓の動きのような elastic に絞って扱っており、それぞれについて形状モデルを事前知識として持つのかか (using a priori shape model), それとも持たないのかで細分している。本稿でメインとなるモデルに基づいた articulated motion のアプローチは、幾何形状モデルの定義, articulated motion モデル, 部位の決定, 三次元構造の決定の四つに分けられている。

さらに、Aggarwal らは人体動作解析に特化したサーベイ^{3),4)}を行っている。ここでは、人体動作解析に関連する主要な研究分野として、(1) 人体動作の解析, (2) 単視点・多視点での人物の追跡^{*}, (3) 人間の活動の認識の三つが挙げられている。人物の姿勢推定は(1)に分類されており、形状モデルを持つものと持たないものがまとめられている。

Cédras ら¹⁸⁾のサーベイは動作に基づく認識に主眼を置いているが、人物動作の追跡と認識という節を割いて、人体動作解析について言及している。そこでは、人体の各部位の認識(ラベリング)と、人体の姿勢や形状の追跡という二つのアプローチが述べられている。

人体と手の動作を扱った Gavrilin⁴⁰⁾のサーベイは、モデルのタイプと追跡する空間の次元という観点に基づいた分類を行っている。形状モデルを持たない二次元アプローチ(姿勢推定することなく画像から動作を認識)、形状モデルを持つ二次元アプローチ(スティックモデルにリボンや blob を加えたもの)、三次元アプローチ(スケルトンのような骨格のみのモデルと、骨格だけでなく肉(flesh)をもつ体積のあるモデル)の三つに分類し、さらに動作の認識について述べている。

Moeslund らのサーベイ⁹⁷⁾は、100以上のモーションキャプチャの文献をそれぞれ詳細に記述したした Technical Report⁹⁶⁾を基に、モーションキャプチャを四つのプロセス——(1)対象の形状モデルと初期姿勢を決定する初期化、(2)人物全体または四肢を背景から抽出しフレーム間の対応をとる追跡、(3)人体形状モデルなどを用いた姿勢の推定、(4)動作の認識——に分割している。主要な部分である(3)では、事

前知識なし(model-free)かモデルを使うのかに分けるだけではなく、使う場合にはさらに、モデルの大きさや位置を参照したり制約として用いる間接的使用(indirect use)と、モデルパラメータを逐次更新していく直接的使用(direct use)に分けている。

Wang ら¹⁷⁰⁾のサーベイは、人体動作解析の主要な三つの問題である(1)人物の検出、(2)人物の追跡、(3)動作の認識を扱っている。(2)の追跡は、形状モデルに基づくもの、領域に基づくもの、動的輪郭モデルを用いるもの、特徴点を追跡するものに分けられている。

Wang ら¹⁶⁹⁾は、追跡、解析と認識という二つの点で分類を行っている。追跡はさらに頭・顔の追跡、手の追跡、人の追跡の三つに分けられ、解析と認識は articulated motion, 全身動作の解析, 三次元姿勢推定, 歩行・ジェスチャ認識の四つに分けられている。

Pentland は、“Looking at People”という視点から関連する技術をサーベイ¹¹⁹⁾しており、個人同定と顔認識、監視とモニタリング、三次元人体追跡、スマートルームと知覚的インターフェース、という四つの分野の state-of-the-art をまとめている。

上記のサーベイはいずれも主に人体の動作を扱っているが、Pavlovic ら¹¹⁷⁾は手のジェスチャ認識に特化したサーベイを行っている。三次元モデルを用いるのか、それとも appearance-base なのかの大きく二つに分類し、ジェスチャのモデル化・解析・認識・応用の四つについてまとめている。解析において三次元形状モデルのパラメータ推定について言及しているが、手の姿勢という特殊性から二次元の形状モデルについては一切触れていない。また彼らは、人体動作解析系のサーベイでは見られない、以下のようなジェスチャ自体のモデル化を示している。ジェスチャ(G)と手の動き(H)と画像(V)の変換を $H = T_{hg}(G)$, $V = T_{vh}(H)$ で表される観測過程と考え、ジェスチャは $V = T_{vh}(T_{hg}(G)) = T_{vg}(G)$ という変換によって観測される。するとジェスチャ認識は $G = T_{vg}^{-1}(V)$ と表現でき、一方姿勢推定は $H = T_{vh}^{-1}(V)$ で表される。

山本^{177),178),179),154)}は動作計測を(1)身体領域の抽出、(2)姿勢の推定、(3)動作の逐次推定の3つのステップに分けた。姿勢の推定は、モデルと画像の重ね合わせによる順運動学と、手先の位置からパラメータを求める逆運動学とに分けられている。また山本¹⁸⁰⁾では動作計測の身体モデル、動作計測、動作認識、感性情報処理についてまとめている。

その他、関連する分野においてグローブを用いた手のモーションキャプチャ¹⁵²⁾や医療分野における歩行解析¹⁷³⁾、CGアニメーション^{161),10),90),39)}のサーベイが行われている。

* ただし全身を一つの物体として追跡するものに限っているため、内容は70)のほうに近い。

3. 画像特徴

モデルフィッティングのための最初の要素は、画像のどのような特徴を用いるのかということである。

人物の三次元形状モデルをトップダウンの知識として与える場合、そのモデルのパラメータを推定するためには観測された画像と何らかの形で比較（マッチングまたはフィッティング）を行う必要がある。当然のことながら、この画像からの観測量（特徴量）は、どのようなモデルを用いてどのような比較を行うのかによって、様々なものが考えられる。しかし、あるモデルを固定してある基準によってフィッティングを行う場合にも、観測量として使える画像の特徴は一つとは限らない。またアプリケーションによっては、ある特徴量は使いやすいが別の特徴量では使用できる条件が限られてしまう、ということも起こり得る。そのため、どのような画像特徴量がよく用いられて、どのような条件において使用できるのかといったことを、手法の構築に際してよく検討し見通しを立てておく必要がある。

この節では、これまでに用いられている画像特徴量を列挙し、代表的な論文においてどのように使用されているのかを示す。また、どのような環境においてどのような特徴が有用であるかについて述べる。表1にそれらの一覧を示す。

3.1 エッジ

モデルを画像にフィッティングする際によく用いられる画像特徴の一つは、エッジである^{29),33),31),41),80),128),127),144),150),150),167),175)}。これは、姿勢が推定された場合には、画像とそこにフィッティングされたモデルのエッジは一致するという考えに基づいている。つまり、

- フィッティングの対象となる人物や手が写っている観測画像のエッジ
- 形状モデルを観測画像と同じ画像平面座標に投影して生成される画像のエッジ

の二つを比較することになる。

後者の画像は、二次元の矩形モデルなどの場合は画像に矩形を描画し、三次元モデルの場合はCGによるレンダリングなどを用いて生成することになる。どちらにしても、描画される画像の背景と、そこに投影されるモデルのとの間には、明確なエッジを検出できるように十分な輝度差があるようにする。またモデルの自己遮蔽（例えば胴体の前面に腕が存在する場合など）に対応するためには、各部位に異なる輝度や色を与えて、投影された部位間にも輝度差があるようにしなければならない。

こうしてできた二つの画像のエッジを比較することになるが、これにはいくつかのバリエーションがある。

- 一つ目は、エッジとしての値を直接比較するやり

方である。

Wachter ら¹⁶⁷⁾ は、モデルを投影したインパルス関数的なエッジに gaussian をかけて滑らかにしたモデルエッジと、観測画像から抽出したエッジとを用いて、各点においてエッジの値の差を計算している。

- 二つ目は、エッジの存在の有無を評価する方法である。

Deutscher ら³¹⁾ は、投影モデルの輪郭上にサンプルした点においてはエッジが存在する値1を設定し、観測画像のエッジを gaussian で滑らかにして0.0~1.0に正規化して、サンプル点での値の差（つまり1-エッジ強度）を評価している。

- 三つ目は、エッジまでの距離を評価するというものである。

Gavrila ら⁴¹⁾ は、モデルを投影したエッジ点からもっとも近い画像中の点までの距離を chamfer matching により評価している。

Wu ら¹⁷⁵⁾ らは、投影したモデルの輪郭上のサンプル点から、法線方向に向かって探索した観測画像におけるエッジ点までの距離を計算している。どの手法にしても、実際の画像のエッジとモデルから生成したエッジとは性質が異なるため、二つをうまく比較する必要がある。

観測画像においては、どのようなシーンが観測されるかによって検出されるエッジの種類が異なり、対策が必要な場合がある。

- ブルーバックや黒一色の暗幕などを用いて、背景が一樣であることを仮定できる場合には、対象となる人物や手のエッジを検出することができる。
- さらに人物の服にテクスチャがない場合や、手や指の場合には、人物と背景とのエッジ、もしくは人物の部位間のエッジに限定される。
- しかし一樣な背景を仮定できない場合、観測対象以外の背景に存在するエッジも検出されてしまうため、これらを排除する必要がある。

背景差分が使用できる場合には対象人物（とそのエッジ）のみを抽出することもできる。背景差分を用いず一樣ではない背景に対処するため、Gavrila ら⁴¹⁾ は、観測画像から抽出したエッジを曲率によってセグメントに分割して分類することで背景のエッジを除去し、人物に対応するエッジだけを残している。

一方、すでに投影モデルのエッジと画像のエッジが近いということが仮定できる場合には、投影モデルの周辺のエッジだけを探索するというやり方もある。Stenger ら¹⁵⁰⁾ は、観測画像のエッジの方向ベクトルに対して、モデルを投影した輪郭の法線ベクトルとの内積をとることで、投影したモデルの法線方向に変化するエッジを用いている。Rehg ら^{127),128)} は、まず円筒モデルを画像に投

表1 画像特徴と環境制約 (○:使用可, △:使用可(ただし計算量が高い, ロバストでない, など), ×:対応不可, ☆:必須)

環境制約	エッジ	シルエット	輪郭線	時空間 勾配	興行き	部位個 別特徴	テンプ レート	領域分割	特徴点
一様背景	○	○	○	○	○	○	○	○	○
非一様背景	△	△	△	△	○	△	△	○	△
静止背景	○	○	○	○	○	○	○	○	○
非静止背景	○	△	△	△	○	○	○	○	△
カメラ固定	○	△	△	○	○	○	○	○	○
カメラ移動	○	×	×	△	○	○	○	○	○
単眼カメラ	○	○	○	○	×	△	○	○	○
複数カメラ	△	○	○	○	☆				
静止画像	○	△	△	×	○	○	○	○/×	○
動画	○	○	○	☆	○	○	○	△/○	○
自己遮蔽有	△	×	×	○	○	△	△	△	△
初期姿勢無	△	△	△	×	△	○	○	○	○
テクスチャ のある服	△	△	△	☆	△	○	☆	△	△
計算量	低	低	低	高	高		高	高	低

影し, その中心線に垂直方向の輝度の微分のピーク(つまりエッジ)を指のシルエットとしている。

3.2 シルエット

シルエットもよく用いられる画像特徴の一つである(21),31),34),68),82),83),98),109),112),140),142),141),175),176)。ここでのシルエットとは, 背景に対して対象物体を前景とみなしたときの前景領域を指す。つまり姿勢が推定された場合には,

- 観測画像中で抽出された人物や手などの対象物体の領域
- 形状モデルを観測画像と同じ画像平面に投影したときにモデルが占める領域

の二つが一致することになる。

後者の投影モデルのシルエット領域は, エッジの場合とは異なり, モデルが存在するかどうかの二値画像を生成すればよい。前者の対象物体のシルエット領域を抽出するには, いくつかの方法がある。

- 背景が一様であることを仮定し, 色や輝度などの二値化や過領域分割により, 前景領域のみを抽出する。
- 参照背景画像などを用いて背景差分を行い, 移動物体としての対象の領域を抽出する。

通常は対象物体の抽出は手法の主要部分ではないため, アプリケーションに応じた仮定を立てて前景領域を容易に抽出できるような環境を整えておく必要がある。そのため, 移動カメラや照明変化が激しいような環境は除外される。

エッジの場合とは異なり, 画像から得られたシルエット領域も投影モデルのシルエットも, どちらも二値のマスクとして同等に扱うことができる。これらのシルエットが一致しているかどうかを比較するには, モデルや対象の形状などを考慮することにより, さまざまな方法が考えられている。Ohyaら¹⁰⁹⁾は, モデルの投影シルエット A と対象のシルエット B の類似度を

$(A \cap B \text{ の面積}) / (A \cup B \text{ の面積})$ で定義し, これを最大化している。亀田ら⁶⁸⁾は, 対象のシルエット S がモデルの投影シルエット C と重なっていない領域(未被覆領域) $S' = S \cap \bar{C}$ の面積を最小にしている。Wuら¹⁷⁵⁾はモデルの投影シルエット A_M と対象の領域 A_I との XOR, つまり $(A_I - A_M)^2$ を評価している。島田ら^{140),142)}は, 対象となる手の指のシルエットから, 投影モデルのシルエットがどのくらい「突出して」はみ出しているかという平均はみ出し長さを評価し, これを小さくしている。Changら²¹⁾は, 画像中の手のシルエット H_I と 3D モデルのシルエット H_R の差分領域を最小化しているが, $H_R - (H_I \cap H_R)$ を小さくするために手を握る shrink force と, $H_I - (H_I \cap H_R)$ を小さくするために手を広げる stretch force を別々に定義している。

シルエットはその性質上自己遮蔽を扱うことができない。胴体の前に上腕と下腕が重なってしまうと, 胴体のシルエットの中に腕が含まれてしまい, 腕の姿勢を判別することができなくなってしまう。そのため, 対象を自己遮蔽のない(両手を広げるなどの)姿勢に限定するか, シルエット以外の特徴を併用することが多くみられる。

3.3 輪郭線

シルエットは面積を持つ特徴であるのに対し, 輪郭線(contour, outline)^{1),2),45),52),155),160)}は背景とシルエットとの境界線としての特徴を持つ。輪郭の抽出の仕方や特徴などはシルエットとほぼ同様であるが, 性質はシルエットよりもエッジに近い。輪郭線がエッジと異なる点は, 対象の外輪郭のみを反映しており, 服のテクスチャや自己遮蔽などによって生じる内部輪郭を持たないということである。

このことを利用して, 両手や両足を広げた人物の輪郭線から, 頭や手先, 足先が輪郭線のどの位置に存在するのかが検出することができる。高橋ら¹⁵⁵⁾は, 輪

軌線上の点から重心と参照位置までの長さ L_t を定義し、その極値を頭部や手先などの位置として検出している。Haritaoglu ら⁴⁵⁾ は対象の輪郭線から convex-hull アルゴリズムにより頭や手足の位置を検出して、立つ、座る、寝転ぶなどの二次元の姿勢推定を行っている。

上述のように、輪郭線は三次元姿勢を推定するにはかなり情報が少なく扱いにくいいため、姿勢推定としてではなく姿勢の認識¹⁶⁶⁾ に用いられることが多い。しかし、Agarwal ら^{2),1)} は、輪郭点を log-polar 形式でヒストグラムを作り、これを形状特徴として SVM や ridge regression, Relevance Vector Machine などで学習し、三次元の姿勢が推定できることを示している。また今井ら⁵²⁾ は、輪郭線上の点により固有空間を生成し、輪郭形状の変化による姿勢推定を行っている。このように、輪郭線は、形状の認識の延長上にある learning-based や example-based の姿勢推定として用いられる画像特徴になるかもしれない。

3.4 輝度および時空間勾配

時空間勾配とは、画像の輝度の時間的変化量（フレーム間の差分）・空間的変化量（エッジ）を指す^{15),19),64),88),101),108),138),182),181)}。

与えられた時系列画像のあるフレーム t において、平面や円筒などの三次元形状モデルの投影がすでに画像中の対象と一致している場合、モデル投影領域上の各点 p_t の次のフレームにおける位置 p_{t+1} もしくは速度 $v(p)$ は、その点における輝度の空間勾配 $\nabla I(p)$ および時間勾配 $I_t(p)$ とモデルの形状から計算することができる。

形状モデルが平面などの剛体の場合、これはよく知られた輝度差を最小にする運動パラメータ推定問題と同じである。しかし、動画像符号化などの運動パラメータによる領域分割や運動物体の個数の推定などは異なり、画像中の形状モデルの投影領域は既知である。人体や手を表す多関節形状モデルの一つ一つの部位のシルエットを作成する際に、背景とシルエットという二値ではなくそれぞれの部位に別々のラベルを割り当てることで、一つの剛体運動パラメータを適用する画像中の領域は限定される。自己遮蔽が生じる場合でも、適切な陰面消去などを行ってシルエットを作成すれば、推定に与える影響は少ない。また単一の剛体物体の運動パラメータの推定ではなく、関節物体の部位の親子関係を制約条件として与えたパラメータの推定が行われている。

前述した画像特徴と時空間勾配が異なる点は、初期値を必要とするということである。エッジやシルエットなどは画像が一枚与えられれば計算でき（シルエット抽出のために参照背景画像は必要ではあるが）、時系列画像では各フレームにおいて別個に処理することができる。しかし時空間勾配を用いる場合は、モデルの投影と画像注の対象が一致するように姿勢パラメータ

の初期値を必要とする。もし一致していない場合には、その一致していない部分において運動パラメータの推定に大きな影響を与える。

陽に時空間勾配を用いて定式化しているものとしては、三次元モデルの二次元運動¹⁸²⁾ や三次元運動^{88),176),15)} を扱ったもの、三次元の平面モデルの運動⁶⁴⁾ に限ったものなどがある。輝度差を最小にする定式化を用いたもの^{101),19)} もある。

時空間勾配を用いた姿勢の追跡の問題点の一つは、各時刻において逐次運動パラメータを推定するため、パラメータに誤差が容易に蓄積してしまう。そのため、パラメータのとり得る範囲に制約を設けること¹⁰⁸⁾ や、推定したパラメータを修正すること¹⁸¹⁾ が提案されている。

ちなみに、オプティカルフローは理論的には時空間勾配と等価な画像特徴であるが、物体の形状を仮定しない通常のフローを精度よく求めることは困難であり、画素毎に求めたフローが運動パラメータ推定自体や姿勢推定に用いられることはあまりない。

3.5 複数カメラの利用

どのような画像特徴を用いても、単眼カメラで複雑な姿勢を観測する場合には、自己遮蔽の問題が生じる。また透視投影を仮定しても、光軸（前後）方向の運動は単眼では判別しにくく、推定に失敗することも多い。これらを解決するために、複数のカメラの情報を用いる手法が提案されている^{28),56),71),107),122),121)15),29),41),138)}。

複数視点の情報を用いるにはいくつかの種類がある。

● ステレオ視による奥行き復元

二枚以上の画像からステレオ視で視差マップ（奥行きデータ）を計算し、それに三次元モデルをフィッティングする手法である。対象人物や手の理想的な奥行き（三次元座標）が得られれば、原理的には自己遮蔽は発生しない。しかし実際には得られる奥行きデータにはノイズが多く含まれている場合があり、密なデータが得られない場合もある。また得られたとしても記憶容量や計算コストが高くなるという問題点もある。

Plänklers ら^{121),122)} らはステレオ視で復元した視差マップを用いて、三次元の二次形式距離を定義し、三次元点とモデルの距離を最小にしている。また Kehl ら⁷¹⁾ は、視野交差で再構成した三次元データと、三次元モデルを構成する表面パッチの各頂点への最短距離を最小にしている。

一方 Delamarre ら²⁸⁾ は、フィッティングに距離を用いるのではなく、三次元モデルと奥行きデータとの間に force を定義してモデルを動かす Maxwell demons¹⁶⁰⁾ という手法を用いている。

- 各視点における定式化をそのまま多視点に拡張
単眼視において最小化問題をそのまま多視点に拡張する手法がある場合、各視点 i での距離の総和 $\sum_j d_i(j)$

のさらに和 $\sum_i \sum_j d_i(j)$ をとる、もしくは(同等であるが)各視点での連立方程式をさらに連立する、など最小化する関数をあまり変更することなく拡張している手法である。

奥行きを保存する必要がないため記憶容量の問題等はないが、最適化問題としての規模が視点数分だけ大きくなるため、単眼視での計算コストの単純和より計算コストはやや増えることになる。

Bregler ら¹⁵⁾ や佐藤ら¹³⁸⁾ は、まず各視点での最小二乗推定のための連立方程式を導き、同期した複数カメラに対してそれらの連立方程式をさらに連立している。

Gavrila ら⁴¹⁾ は、各視点でそれぞれの類似度を計算し、それらを基にすべての視点における単一の類似度を評価している。

● 各視点における結果を多視点で統合

単眼視において処理された結果がある場合に、それらを統合処理する方法である。例えばそれぞれの視点においてベクトルを算出した後に、三次元的なベクトルの和をとる、などがある。

Delamarre ら²⁹⁾ は、各視点において投影した三次元モデルの輪郭と対象のシルエットとの間に force ベクトルを定義している。それらのベクトルを和をとって三次元モデルを動かす force を計算し、モデルをフィッティングしている。

3.6 部位個別特徴

人体は各部位毎にそれぞれ色やテクスチャや形状などの性質が違っているため、全ての部位に同じ画像特徴を用いるのではなく、部位の特徴に応じて異なった画像特徴を用いる手法もある^{49),74),84),186)}。各部位が別個に処理されるため、そのまま最終結果とするのではなく、全体の最適化における探索範囲を限定したり、統一的な尤度計算に用いたりして、全体としての整合を取ることが必要となる。

北嶋ら⁷⁴⁾ は頭部は近似楕円との領域の重なりで、胴体は矩形内の画素値で、二直線でモデル化した腕は輪郭線で、それぞれの部位を検出し、GA による最適化の探索空間を絞り込んでいる。

Lee ら⁸⁴⁾ は Data-driven MCMC (Markov Chain Monte Carlo) を使ったアプローチにおいて、部位毎に異なった画像観測量を尤度計算に用いている(顔は adaboost による検出、頭と肩の輪郭線は動的輪郭モデル、顔、両手、両足は肌色楕円抽出、両足は中心線)。また Hua ら⁴⁹⁾ も同様に、data driven belief propagation Monte Carlo (DDBPMC) という提案手法において、部位毎に異なった画像観測量を用いて重点的サンプリングに使用している(顔は ada-boost 顔検出、下腕下肢は顔と肌色検出と矩形の当てはめ、胴体はエッジ線分の抽出)。

3.7 テンプレート

各部位の見えをテンプレート等によって学習し、そ

れによって各部位を検出して全体の姿勢を推定する手法もある^{123),124)}。部位を別々に検出すると全体として整合が取れなくなるため、それぞれの部位間の接続関係などを抽出の制約にする必要がある。

Ramanan ら^{123),124)} は pictorial structure^{36),35)} に基づく手法により、各部位を矩形で表した見えモデルを用いて各フレームで部位を検出し、モデル間の接続関係を kinematics としてあらかじめ与えて、Bayesian net で姿勢を推定している。

3.8 領域分割

色や運動に基づく領域分割を用いて、一般の背景の中に存在する対象の部位を領域として検出して姿勢推定を行うという試みもある^{77),76),100),156)}。

静止画像については、Mori ら¹⁰⁰⁾ は canny edge や normalized cuts を用いて境界線に基づいて分割した候補領域の中から、四肢や胴体を検出した後、各部位のつながりを認識して姿勢推定を行っている。

動画像については、Krahnstoever ら^{77),76)} は、EM アルゴリズムによる運動パラメータ推定と領域分割を行い、単眼での上腕と下腕の三次元円筒モデルの関節角の推定を行っている。また Tamaki ら¹⁵⁶⁾ はオプティカルフローを用いた領域分割と運動パラメータ推定を行い、関節位置を推定している。

3.9 特徴点

コーナーなどの特徴点を検出し、それを用いて姿勢推定を行う手法もある^{46),57)}。

岩井ら⁵⁷⁾ は対象となる手の輪郭からコーナーを検出し、指先と指の又(付け根)の位置を特徴点として検出して姿勢推定を行っている。

林ら⁴⁶⁾ は対象形状の長さの知識を利用してキャリブレーションを行いながら姿勢推定を行う Calibration Using target Knowledge (CUK) 手法を提案し、Harris 式コーナー検出器で腕の両端(拳の先と肘)を検出している。

3.10 対応点

多関節物体の姿勢や関節位置を推定するために、与えられた対応点から剛体の形状とカメラの運動を推定する shape-from-motion の研究を発展させた手法もある^{162),163),187),188)}。幾何モデルは用いていないが、ここで述べておく。

Zhang ら^{187),188)} は、与えられた対応点から関節位置と関節角度を推定しており、 N 個の部位がある場合には各部位について $5N (> 3)$ 点あれば初期姿勢が決定でき、追跡のためには全体で $N + 5$ 点あればよいことを示し、ロボットアームで実験し真値と比較評価している。

Tresadern ら^{162),163)} は、与えられた対応点から link への分割と joint の位置を求める、non-rigid, uncalibrated, unsynchronized な多視点映像の因子分解法を提案し、ユニバーサルジョイントとヒンジで実験を行っている。

4. モデルの表現とパラメータ空間

画像特徴の次のモデルフィッティングのための要素は、どのようなモデルを用いるのかということである。

関節物体は剛体と見なすことができる「部位」と部位同士を結合する「関節」からなっている。このことから、関節物体のモデリングする場合は、部位を関節によって結んだ多関節モデルが通常用いられることになる。

4.1 モデルの位置姿勢パラメータ

モデルフィッティングの目的は対象の関節物体の位置姿勢を求めることなので、モデルも位置姿勢のパラメータを持っていなければならず、モデルと画像特徴を照合することによってこのパラメータを求めることが可能なことがモデルフィッティングの最低条件である。

モデルの位置姿勢パラメータの表現方法としては以下のようなものが良く使われる。

4.1.1 各部位の位置姿勢

各部位ごとの位置姿勢をモデル全体のパラメータとして利用する方法である。各部位は関節によって結ばれているため、それらの位置姿勢は独立ではないことから、パラメータには冗長性が存在する。この冗長性を取り除くために、パラメータ推定時に制約を加えなければならない。

4.1.2 各関節および各端点の位置

各関節点および各端点の位置をモデル全体のパラメータとして利用する方法である。関節および端点の間の距離は部位の長さによって決まっているため、それらの位置は独立ではないことから、パラメータ推定には冗長性が存在する。この冗長性を取り除くために、パラメータ推定時に制約を加えなければならない。さらに、部位のロール方向の回転を表すことができないという問題も存在する。

4.1.3 各関節の角度と基準となる部位の位置姿勢

部位間の相対的な位置姿勢を表す各関節の角度と、関節物体全体の位置姿勢を表す基準となる部位の位置姿勢をモデル全体のパラメータとして利用する方法である。他の表現方法と異なり、パラメータに冗長性は存在しない。

4.2 部 位

この節では、これまでに提案されている関節物体モデルにおける部位の表現法を列挙し、代表的な論文においてどのように使用されているのかを示す。また、それぞれの特徴についても述べる（表 2 参照）。

4.2.1 線 分

部位の表現によく用いられるものとして線分^{72),69),107),183),11),22),86),24)}が挙げられる。これは、部位を表現するために最低限の情報のみからなる非常にシンプルな表現であるため、高速に計算を行うことができ、

モデルの構築が容易である。一方で、部位の断面方向の情報やテクスチャ情報を持たせることができないため、断面方向のあいまい性が大きく正確な位置を求めることが難しいほか、部位のロール方向の回転を扱うことが不可能であるという問題がある。

4.2.2 平面領域

各部位の形状を平面領域で表現する手法も提案されている。画像中の部位の形状を近似するような平面領域を用い、それと画像特徴と照合することによって、画像中の部位の位置を推定することができる。

平面領域としては、四角形^{175),85),64),87)}、楕円⁷⁵⁾、リボン⁸⁰⁾、凸包⁸⁷⁾、プロップ⁷⁷⁾、テンプレート画像¹²⁶⁾などが一般的に用いられている。

例えば、Lanら⁸²⁾は、人物モデルとして pictorial structure を利用している。pictorial structure とは複数の部位とそれを結合するばねからなるモデルであり、各部位はその部位のある視点からの外観を表すパラメータ、各ばねは部位どうしの相対的な距離や方向とその共分散（ねじれや伸びの大きさの程度を評価する）を表すパラメータを保持している。さらに姿勢推定に HMM を導入している。HMM の各状態が pictorial structure model のある状態（代表的な姿勢）に対応しており、これによって姿勢推定を行っている。視点方向の変化も HMM の枠組みで扱っている（つまり、代表的な視点位置からの代表的な姿勢が HMM の各状態に対応）。モデルのパラメータおよび HMM の遷移確率は、例題からの学習によって獲得している。代表的な姿勢の数が限られる場合（実験では人物の歩行を対象としている）に有効な手法である。

4.2.3 立 体

体積を持つ立体によって部位を表現する手法も挙げられる。関節物体は立体であるので、立体によって部位を表現することにより、その部位の形状特徴をより正確に表現することができる。また、モデルが三次元表現であるため、多視点画像や距離画像を用いた手法との相性も良い。一方で、モデルパラメータが多くなるため計算量は増大してしまう。

立体としては、直方体¹⁴⁰⁾、円筒⁷⁶⁾、円錐台^{167),30),31),143)}、三次元プロップ⁶³⁾、二次曲面¹⁵⁰⁾、超二次曲面^{41),66),106),91)}などが用いられている。

さらに詳細なモデルとして、CG 分野で利用されている多面体モデルを利用する手法もある^{68),164),33),191),21),71),27)}

モデルの構築は最も困難であるが、それを基に作成する合成画像の正確さに加え、グラフィックスハードウェアの高速化もあって、Analysis-by-Synthesis の枠組みによるモデルフィッティングにおいては強みを発揮する。Plänkersら¹²²⁾は、モデルとして CG 分野で利用されているメタボールを関節でつなげたものを用いている。3眼ステレオによる距離画像とシルエットから各部位の位置姿勢を求めている。また、山本ら

表 2 部位の表現方法

表現方法	画像特徴				部位パラメータ数	モデル構築
	エッジ	シルエット	時空間勾配	奥行き		
線分	難	適	不可	難	少	易
平面領域	適	適	適	中	中	中
立体	適	適	適	適	多	難

のグループ^{182),176),136),108)}では、CADモデルを用いて、時空間勾配法で姿勢推定を行っている。

4.3 関節

部位と並ぶモデルのもうひとつの要素は関節である。関節は部位と部位を接続する働きを持ち、通常は部位中の1点として表される。つまり、ある部位の1点と隣の部位の一点とが同じ位置にあることになる。

部位同士を関節の1点によって結ぶのではなく、部位同士をばねのようにゆるく結ぶ手法も考えられる^{82),174),118),143)}。こうすることで、部位形状の正確さをばねの伸び縮みで吸収できるが、パラメータ数は大幅に増加する。

4.4 その他のモデル表現方法

これまでは部位と関節との組み合わせで関節物体を表現する方法について紹介してきたが、このような表現以外のモデル表現方法も提案されている。

島田らのグループ^{44),52)}では、手の輪郭をモデルとしている。輪郭を等間隔に標準化し、標準点の座標を形状特徴としている。すべての姿勢について形状特徴を用意しておき、入力の特徴と比較することで、手指の姿勢を求めている。

Zhangら¹⁸⁶⁾は、非剛体の二次元身体形状の輪郭を表現するため Bayesian graphical model を提案している。Bayesian graphical model とは T 個のランドマーク (部位ごとの輪郭上の点) をノードとするベイジネットである。ランドマークどうしを結んで三角形を作り、ネットワークを構成している。さらに9個の関節角度どうしの関連を表すために、9個の関節をノードとするベイジネットも利用している。この場合、はすべてのランドマークの二次元位置がパラメータとなる。これらのベイジネットを利用して確率的に姿勢推定を行っている。

4.5 モデルの学習

モデルフィッティングでりようするモデルには、各部位の形状パラメータのように、数多くのパラメータが存在する。これまでの多くの研究ではこれらのパラメータはシステム開発者や利用者によって事前に与えられていた。しかし、これらのパラメータを例題画像から自動的に獲得することを目指す研究も行われている。

Lan らの研究⁸²⁾では、pictorial structure のパラメータ、部位どうしの位置関係を例題画像から学習している。

Krahnstoever らは、単眼画像におけるモーション

セグメンテーションによって部位であるブロップのパラメータを学習している⁷⁷⁾。さらに同じグループでは部位として可変円筒を用いて、そのパラメータを同様に学習している⁷⁶⁾。

部位の形状に関するパラメータを事前に学習するだけでなく、モデルフィッティング時に位置姿勢パラメータの推定と同時に形状に関するパラメータの推定も行う研究もある。パラメータ数がさらに増えることになるのでパラメータ推定は難しくなる。

Marzani らの研究⁹¹⁾や Nunomaki らの研究¹⁰⁶⁾では、部位として用いられている変形可能 (前者では先細り、後者では先細りと曲げが可能) な超二次曲面のパラメータを位置姿勢パラメータと同時に推定している。

前述した Zhang らの研究¹⁸⁶⁾では、ランドマークの画像特徴、ランドマークどうしの位置関係、9個の関節角度どうしの関連を例題画像から学習している。さらに、学習によって得られたこれらのパラメータの変動幅を基に、モデルマッチングにおけるランドマークの位置ずれ (つまり部位の変形) を行っている。

5. 最適解の探索手法

5.1 関節物体の最適解探索問題

一般にモデルを画像に当てはめる問題は、以下の段階にわけて定義される。

- モデル (パラメータ θ) を定義する。 θ を変化させたときに画像と照合するべき特徴 $F_m(\theta)$ を生成する方法もあわせて定義する。
- 対応するモデル特徴 $F_m(\theta)$ と画像特徴 F_i の照合度 $S(F_m(\theta), F_i)$ (評価関数) を定義する
- 評価関数 S を最大化 (あるいは最小化) する解 $\hat{\theta}$ をなんらかの方法に基づいて求める

評価関数は、3節で詳述した特徴ごとに、特徴どうしの類似性 (対応するモデル、画像特徴間の画像上の距離、エッジの方向、色分布、テクスチャなど) に基づいて決められる。モデルが三次元剛体の場合、物体の位置と回転の6自由度のパラメータなどについて、最適解を探索する問題となる。

一方、関節物体に代表される物体内部に変形の自由度をもった物体については、6自由度に内部自由度を加えたパラメータ空間において最適解を探索する問題となるが、探索すべき自由度が増えると探索空間が指数的に増大し実用上最適解を求めることが不可能とな

る。また、人体のような構造複雑な物体では画像上での自己隠蔽によって解にあいまいさが生じる場合があり、そのような場合に対処できる手法をいかに構成するかが課題となっている。

そこで、物体自体の構造や時系列における姿勢の連続性仮定など、なんらかの拘束条件を活用した探索空間の制限を行うことによって、効率よくかつ正しい解を（場合によっては複数）見つける方策がさまざまに提案されている。

5.2 標準的な数値解法に基づく最適化

画像特徴として、人体であれば頭部、手先の位置や腕、胴体の輪郭線がシルエットなどから十分よく抽出でき、モデルのどの部分に対応するかが正しく決定できる場合には、評価関数 $S(\theta)$ を、モデル特徴と画像特徴の差異の二乗和として定義でき、いわゆる最小二乗法の枠組みによって最適解の探索を行うことができる。

多くの場合、剛体回転自由度に起因して $S(\theta)$ の非線形性は非常に強いため、一般的な評価関数の最適化手法である最急降下法、ニュートン法などを直接用いると収束性が悪い、Jacobian や Hessian の計算が解析的に困難などの理由によりあまり用いられない。時系列追跡における連続性仮定によって十分解の近くに初期解を配置できれば、ガウス・ニュートン法やその改良法であるレーベンバーグ・マーカート法などの簡便で性能のよい探索手法を用いることで、比較的高速に妥当な解が見つかる^{88),127)}。これら手法では前時刻での姿勢推定に基づいてモデルと対応付けすべき画像特徴の探索範囲を限定し、物体の同一部位をロバストな追跡を試みている。

ただし、いずれの場合も局所解問題を解決することは一般に難しく、距離画像やマルチビューによる 3D の情報に基づいて姿勢パラメータを推定する場合には、ICP (iterative closest point) 法によるモデル-画像間対応付けと最小化を繰り返す手法によって、2D の投影画像に当てはめる場合よりも比較的良好な解が得られている²⁸⁾。

また、特定部位の特徴を単独で追跡するのではなく、時系列画像から各ピクセルのオプティカルフローによる運動情報を画像特徴として用いることで、各剛体パーツの微小三次元運動を復元し、剛体部位全体の追跡を行う手法がある^{15),176),138)}。この場合、時間とともにモデルの部位と画像中の対応部位が累積誤差によってしだいにずれていく問題（ドリフト）があり、運動の最初と最後のフレームにおいて正解を与えることによってドリフトを修正する方法が提案されている¹⁸¹⁾。

5.3 探索空間の削減

多関節物体のような大きな自由度をもつモデルパラメータ空間は膨大なため、対象物体に関する構造上の拘束をつかって探索空間を限定することが行われる。研究の初期には、関節物体は根元の姿勢が末端の姿勢

に影響を与えることに着目し、根元から末端に向かって順番にパラメータの最適化を行う方法^{95),78),68),140)}が多く試みられたが、根元の部位の姿勢を単独で一意に決定できることが少なく、バックトラックなどの工夫が必要となる。

近年では、人体のような各関節の動きに相関が見られる場合を対象に、多量のサンプルデータから次元圧縮を行い、探索自由度を減らすことが行われている^{111),189),175)}。どの程度自由度を圧縮できるかはどのような動作を対象とするかに強く依存するが、三次元的な姿勢の推定を目的とする場合、各部位の追跡が困難にならない程度に抑える必要がある。

現時刻の画像情報に加えて、過去の情報と何らかのダイナミクスに基づく予測を行い、探索範囲を限定する方法は有効であり、著名なものにカルマンフィルタ⁷⁾がある。評価関数 S は一般に非線形なので Jacobian が計算できる場合には拡張カルマンフィルタ (EKF) が用いられる^{142),92)}。カルマンフィルタにおける共分散行列は不確かさを表現しているので、対応する特徴を探索する際の探索範囲の設定基準に示唆を与える。

EKF では線形化による近似誤差のため、quaternion などの回転変換系や透視変換を含む非線形系ではバイアスを生じて追跡失敗を起こしやすい。そこで、EKF の平均値の計算時に、共分散楕円上のサンプル点の情報を用いてバイアスを補正する Unscented Kalman Filter (UKF)⁶⁵⁾ が近年用いられている^{168),149)}。

他の予測手法としては、自己回帰モデルを用いるもの⁵⁵⁾ や音声認識の代表的なモデリング手法である隠れマルコフモデル (HMM) やその拡張を用いるもの^{14),60)} がある。

5.4 姿勢推定の曖昧性への対処

画像によるモデル当てはめでは、一部の部位が自己隠蔽によって観測できない、また異なる姿勢パラメータからよく似た見えが生成され、一時的にはどちらか判別ができないといった姿勢推定の曖昧性の問題がある。単一の状態だけ保持して追跡を続けることは本質的に困難であり、複数の候補を並列に保持しながら追跡を行う手法が現在では一般的である。これは多峰性分布における確率的状態推定問題に相当しており、背景からの特徴切り出しなど入力に誤りを含む場合の最適解探索にも効果的である。具体的には、古典的な beam-search に基づく並列探索方法¹⁴⁰⁾ や、particle filter、CONDENSATION といったモンテカルロ手法に基づく手法^{55),151),175)} が用いられている。モンテカルロ法ではパラメータ自由度の増加とともに必要とされるサンプルの数が指数関数的に増大するため、前節で述べた探索空間の削減手法とともに適用されることがほとんどである。

事前に用意された generic なモデルの形状（長さ、太さといった剛体部分のパラメータ）を入力にあわせて適応的に変化させることも行われているが、画像の

奥行き曖昧性のために、その瞬間の画像だけではそもそも一意に決定できない場合がある。そのような場合に、各パラメータのとりうる可能な範囲やパラメータ同士の関係性を知識として導入し、整合性のあるパラメータの範囲を領域として追い込むことによってパラメータを限定する方法がある^{113),142)}。また画面内の密な動き情報をつかって画像から得られる制約をリッチにすることでモデル形状を一意に決定する方法がある⁸⁹⁾。

5.5 見えの学習に基づく探索

一般の背景のもとでは、物体各部に対応付けできる明確な画像特徴を安定的に抽出することは難しいので、全体の見え(シルエット、テクスチャやHu-momentなどの特徴)をモデルから次々に生成しておき、パラメータの変化とそれに伴って現れる見えの対応をなんらかの表現方法で記述しておく方法が近年主流になっている。

直接的な方法は入力特徴からモデルパラメータへの写像関数を、モデルから生成したサンプルをもとに学習しておき、この写像関数を用いて実画像の入力特徴からモデルパラメータを計算する方法である。この写像関数は極めて非線形性が強いので、区分的な関数の集合や局所的なカーネル関数の級数として表現する方法で実現される。¹³⁴⁾は mixture gaussian 分布のパラメータを MAP 推定することで写像関数を決定している。⁴³⁾では自己組織化マップ(Self-Organizing Map: SOM)¹⁵⁸⁾を用いて、よく似た写像関係を与える入力特徴-パラメータの組を自律的に集め、それぞれに最適化された写像関数を決定している。

一方、写像を関数として表現するのではなく、特徴-パラメータ空間における多様体として定義し、非線形性の強い多様体を小さいパッチに分割して記述することも行われている⁴⁷⁾。入力がどのパッチに属しているのかは特徴空間における入力と最近傍パッチで与えられ、各パッチとの距離(もしくは距離に基づいて計算される確率密度)が照合度を表すので、CONDENSATION などモンテカルロ法による多峰性分布の推定および追跡を行うことができる。

6. 歴史と研究動向

近代的な人体動作の観察は 1870 年代の Eadweard Muybridge^{102),103)} や Etienne-Jules Marey にまでさかのぼるだろう。その後も人体運動の知覚⁹⁹⁾ という心理学的な研究テーマとなり、Moving Light Display (MLD) による biological motion^{62),61)} などの研究が盛んに行われるようになり、三次元復元の研究^{171),172)} も行われた。

1970 年代から CG の分野ではすでに人体動作を表現するための研究が行われていた¹⁰⁾ が、コンピュータビジョンの分野において人体のモデルに基づいて動

作を解析しようとする試みは、1980 年代の O'Rourke¹¹³⁾ や Hogg⁴⁸⁾ などが最初である。

1990 年代になると、計算機環境の高性能化と低価格化、さらに動作計測へのニーズの高まり¹⁷⁸⁾ などにより、姿勢推定は活発な研究分野となった。1980 年代に研究された rigid motion estimation の流れから、90 年代初頭には Lowe⁸⁸⁾ や Yamamoto¹⁸²⁾ の関節物体の追跡手法が提案され、その後も Bregler¹⁵⁾ や Ju⁶⁴⁾ など、90 年代を通じて運動パラメータ推定に基づく姿勢の追跡が研究対象となる。モデルを画像に投影してマッチングを行う analysis-by-synthesis (または generate-and-test) strategy は、O'Rourke¹¹³⁾ が prediction という考え方を示していたが、1990 年代半ばに複数カメラで複数の人物の姿勢を追跡してみせた Gavrila⁴¹⁾ の影響が大きい。また、人体や手は単に non-rigid ではなく articulated なモデルであるということを明確に示した Aggarwal⁵⁾ に始まる多くのサーベイが、この分野の研究の発展を促してきた。90 年代は、ICCV, CVPR, ECCV, ACCV などの国際会議のみならず、IEEE International Conference on Automatic Face and Gesture Recognition (FG) や IEEE Computer Society Workshop on Motion of Non-Rigid and Articulated Objects (NAM), IEEE Computer Society Workshop on Human Motion (HUMO) などのこの分野に特化したワークショップが開催されるなど、大きな進展を見せた時期であると言えよう。

2000 年以降になると、見えと姿勢の関係を学習する learning-base^{52),2),1)} の手法や、人体の部位間の接続が回転のみである(硬い) articulated モデルではなく部位間の移動もある程度許すような(緩い)人体モデル¹⁴³⁾ を用いるもの、人体モデルというトップダウンだけではなく各部位の情報から人体を組み上げるボトムアップ的な手法^{53),35),54),123),84),49)}、structure-from-motion を関節物体に拡張したもの^{162),163),188)} など、いろいろな手法が出現してきており、今後も更なる発展が期待される。

7. おわりに

本稿では人体や手などに代表される多関節物体の三次元姿勢を画像から推定するモデルフィッティングの技術についてサーベイした。画像によるモデルフィッティングの枠組みを、1) 推定に利用される画像特徴、2) 画像と照合するモデルの表現と照合のパラメータ空間、3) 照合時の評価関数と最適解の探索手法、にわけて多関節物体の三次元姿勢推定に特徴的な要素を上記三つの観点から比較整理することを試みた。

なお、情報処理学会のウェブサイトで開催できる本稿の pdf ファイル版には、データベースとしての利便性を提供するため、参考文献欄に本稿で引用した文献

のオンラインリソースへの hyper link を埋め込んで
いるので、読者には活用されたい。

参 考 文 献

- 1) Agarwal, A. and Triggs, B.: 3D Human Pose from Silhouettes by Relevance Vector Regression, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp.882–888 (2004). 3.3, 6
- 2) Agarwal, A. and Triggs, B.: Recovering 3D Human Pose from Monocular Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.28, No.1, pp.44–58 (2006). 3.3, 6
- 3) Aggarwal, J.K. and Cai, Q.: Human Motion Analysis: A Review, *IEEE Workshop on Motion of Non-Rigid and Articulated Objects*, pp. 90–102 (1997). 2
- 4) Aggarwal, J.K. and Cai, Q.: Human Motion Analysis: A Review, *Computer Vision and Image Understanding*, Vol.73, No.3, pp.428–440 (1999). 2
- 5) Aggarwal, J.K., Cai, Q., Liao, W. and Sabata, B.: Articulated and elastic non-rigid motion: A review, *IEEE Workshop on Motion of Non-Rigid and Articulated Objects*, pp.16–22 (1994). 2, 6
- 6) Aggarwal, J.K., Cai, Q., Liao, W. and Sabata, B.: Nonrigid Motion Analysis: Articulated and Elastic Motion, *Computer Vision and Image Understanding*, Vol.70, No.2, pp.142–156 (1998). 2
- 7) 有本 卓:カルマン・フィルタ, 産業図書 (1977). 5.3
- 8) Arman, F. and Aggarwal, J.K.: Model-based object recognition in dense-range images - a review, *ACM Computer Survey*, Vol.25, No.1, pp. 5–43 (1993).
- 9) Athitsos, V. and Sclaroff, S.: Estimating 3D Hand Pose from a Cluttered Image, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol.2, p.432 (2003).
- 10) Badler, N.I. and Smoliar, S.W.: Digital Representations of Human Movement, *ACM Computing Surveys*, Vol.11, No.1, pp.19–38 (1979). 2, 6
- 11) Barrón, C. and Kakadiaris, I.A.: On the improvement of anthropometry and pose estimation from a single uncalibrated image, *Machine Vision and Applications*, Vol.14, No.4, pp.229–236 (2003). 4.2.1
- 12) Ben-Arie, J., Pandit, P. and Rajaram, S.: View-Based Human Activity Recognition by Indexing & Sequencing, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol.2, p.78 (2001).
- 13) Bergen, J.R., Anandan, P., Hanna, K.J. and Hingorani, R.: Hierarchical Model-Based Motion Estimation, *European Conference on Computer Vision*, London, UK, Springer-Verlag, pp.237–252 (1992).
- 14) Brand, M., Oliver, N. and Pentland, A.: Couple Hidden Markov Models for Complex Action Recognition, *IEEE International Conference on Computer Vision and Pattern Recognition*, p.994 (1997). 5.3
- 15) Bregler, C. and Malik, J.: Tracking People with Twists and Exponential Maps, *IEEE International Conference on Computer Vision and Pattern Recognition*, p.8 (1998). 3.4, 3.5, 5.2, 6
- 16) Buxton, H.: Learning and understanding dynamic scene activity: a review, *Image and Vision Computing*, Vol. 21, No. 1, pp. 125–136 (2003).
- 17) Calvert, T.W. and Chapman, A.E.: Analysis and Synthesis of Human Movement, *Handbook of Pattern Recognition and Image Processing: Computer Vision* (Young, T. Y., ed.), Vol. 2, Academic Press, pp.431–474 (1994).
- 18) Cédras, C. and Shah, M.: Motion-based recognition: a survey, *Image and Vision Computing*, Vol.13, No.2, pp.129–155 (1995). 2
- 19) Cham, T.-J. and Rehg, J.M.: A Multiple Hypothesis Approach to Figure Tracking, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol.2, p.2239 (1999). 3.4
- 20) Chang, C., Ansari, R. and Khokhar, A.: Cyclic Articulated Human Motion Tracking by Sequential Ancestral Simulation, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol.2, pp.45–52 (2004).
- 21) Chang, W.-Y., Chen, C.-S. and Hung, Y.-P.: Appearance-Guided Particle Filtering for Articulated Hand Tracking, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. 235–242 (2005). 3.2, 4.2.3
- 22) Cheung, G. K., Baker, S. and Kanade, T.: Shape-From-Silhouette of Articulated Objects and its Use for Human Body Kinematics Estimation and Motion Capture, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. 77–84 (2003). 4.2.1
- 23) Chu, C.-W., Jenkins, O.C. and Mataric, M.J.:

- Markerless Kinematic Model and Motion Capture from Volume Sequences, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol.2, p.475 (2003).
- 24) Date, N., Yoshimoto, H., Arita, D. and Taniguchi, R.: Real-time Human Motion Sensing based on Vision-based Inverse Kinematics for Interactive Applications, *IEEE International Conference on Pattern Recognition*, Vol.3, pp.318–321 (2004). 4.2.1
- 25) de Campos, T. E., Tordoff, B. J. and Murray, D.W.: Linear Recovery of Articulated Pose Change: Comparing Pre- and Post-imposed Constraints, Technical Report OUEL 2279/05, Department of Engineering Science, Oxford University (2005).
- 26) de Campos, T. E., Tordoff, B. J. and Murray, D.W.: Recovering Articulated Pose: A Comparison of Two Pre and Postimposed Constraint Methods, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.28, No.1, pp.163–168 (2006).
- 27) dela Rivière, J.-B. and Guitton, P.: Model-based video tracking for gestural interaction, *Virtual Reality*, Vol.8, No.4, pp.213–221 (2005). 4.2.3
- 28) Delamarre, Q. and Faugeras, O.: Finding Pose of Hand in Video Images: A Stereo-Based Approach, *International Conference on Face & Gesture Recognition*, pp.585–590 (1998). 3.5, 5.2
- 29) Delamarre, Q. and Faugeras, O.: 3D Articulated Models and Multi-View Tracking with Silhouettes, *International Conference on Computer Vision*, Vol.2, p.716 (1999). 3.1, 3.5
- 30) Delamarre, Q. and Faugeras, O.: 3D Articulated Models and Multiview Tracking with Physical Forces, *Computer Vision and Image Understanding*, Vol.81, No.3, pp.328–357 (2001). 4.2.3
- 31) Deutscher, J., Blake, A. and Reid, I.: Articulated Body Motion Capture by Annealed Particle Filtering, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol.2, p.2126 (2000). 3.1, 3.2, 4.2.3
- 32) Deutscher, J., Davison, A. and Reid, I.: Automatic Partitioning of High Dimensional Search Spaces Associated with Articulated Body Motion Capture, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol.2, p.669 (2001).
- 33) Drummond, T. and Cipolla, R.: Real-Time Visual Tracking of Complex Structures, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.24, No.7, pp.932–946 (2002). 3.1, 4.2.3
- 34) Elgammal, A. and Lee, C.-S.: Inferring 3D Body Pose from Silhouettes Using Activity Manifold Learning, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol.2, pp.681–688 (2004). 3.2
- 35) Felzenszwalb, P. F. and Huttenlocher, D. P.: Efficient Matching of Pictorial Structures, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol.2, p.2066 (2000). 3.7, 6
- 36) Fischler, M.A. and Elschlager, R.A.: The Representation and Matching of Pictorial Structures, *IEEE Trans. on Computers*, Vol. 22, No.1, pp.67–92 (1973). 3.7
- 37) Forsyth, D. A. and Ponce, J.: *Computer Vision: A Modern Approach*, Prentice Hall. (2002).
- 38) Forsyth, D. A. and Ponce, J.: Tracking with Non-Linear Dynamic Models, *Chapters that didn't make it into the final version in Computer Vision: A Modern Approach*, pp.33–56 (2002).
- 39) Garchery, S., Boulic, R., Capin, T. and Kalra, P.: Standards for Virtual Human, *Handbook of Virtual Humans*, John Wiley & Sons, Ltd, pp. 373–391 (2004). 2
- 40) Gavrilu, D.M.: The Visual Analysis of Human Movement: A Survey, *Computer Vision and Image Understanding*, Vol. 73, No. 1, pp. 82–98 (1999). 2
- 41) Gavrilu, D.M. and Davis, L.S.: 3-D model-based tracking of humans in action: a multi-view approach, *IEEE International Conference on Computer Vision and Pattern Recognition*, pp.73–80 (1996). 3.1, 3.5, 4.2.3, 6
- 42) Goldgof, D.B., Lee, H. and Huang, T.S.: Motion analysis of nonrigid surfaces, *IEEE International Conference on Computer Vision and Pattern Recognition*, pp.375–380 (1988). 2
- 43) Guan, H., Feris, R.S. and Turk, M.: The Isometric Self-Organizing Map for 3D Hand Pose Estimation, *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 263–268 (2006). 5.5
- 44) 浜田康志, 島田伸敬, 白井良明: 遷移ネットワークに基づく多視点画像時系列からの手指形状推定, 電子情報通信学会論文誌 D-II, Vol.J85-D2, No.8, pp.1291–1299 (2002). 4.4
- 45) Haritaoglu, I., Harwood, D. and Davis, L.S.: W4: Real-Time Surveillance of People and Their Activities, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.22,

- No.8, pp.809-830 (2000). 3.3
- 46) 林健太郎, 久野義徳, 島田伸敬, 白井良明: 動的ロバストキャリブレーションによる人体の姿勢復元, 電子情報通信学会論文誌 D-II, Vol.J83-D2, No.3, pp.977-987 (2000). 3.9
- 47) Heap, T. and Hogg, D.: Wormholes in Shape Space: Tracking Through Discontinuous Changes in Shape, *International Conference on Computer Vision*, pp.344-349 (1998). 5.5
- 48) Hogg, D.: Model-based vision: a program to see a walking person, *Image and Vision Computing*, Vol.1, No.1, pp.5-20 (1983). 6
- 49) Hua, G., Yang, M.-H. and Wu, Y.: Learning to Estimate Human Pose with Data Driven Belief Propagation, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol.2, pp.747-754 (2005). 3.6, 6
- 50) Huang, T.S.: Modeling, analysis, and visualization of nonrigid object motion, *IEEE International Conference on Pattern Recognition*, Vol.1, pp.361-364 (1990). 2
- 51) Humanoid Animation Working Group: H-anim. www.h-anim.org.
- 52) 今井章博, 島田伸敬, 白井良明: 輪郭の変形の学習による 3-D 手指姿勢の認識, 電子情報通信学会論文誌 D-II, Vol.J88-D-II, No.8, pp.1643-1651 (2005). 3.3, 4.4, 6
- 53) Ioffe, S. and Forsyth, D.: Finding People by Sampling, *International Conference on Computer Vision*, Vol.2, p.1092 (1999). 6
- 54) Ioffe, S. and Forsyth, D.: Human Tracking with Mixtures of Trees, *International Conference on Computer Vision*, Vol.1, p.690 (2001). 6
- 55) Isard, M. and Blake, A.: Contour Tracking by Stochastic Propagation of Conditional Density, *European Conference on Computer Vision*, pp.343-356 (1996). 5.3, 5.4
- 56) 石井浩史, 望月研二, 岸野文郎: 人物像合成のためのステレオ画像からの動作認識法, 電子情報通信学会論文誌 D-II, Vol.J76-D2, No.8, pp.1805-1812 (1993). 3.5, 4.2.1
- 57) 岩井儀雄, 八木康史, 谷内田正彦: 単眼動画像からの手の 3 次元運動と位置の推定, 電子情報通信学会論文誌 D-II, Vol.J80-D2, No.1, pp.44-55 (1997). 3.9
- 58) 岩井儀雄: モデルベース手法による身体動作計測, 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研究報告, Vol.2001, No.066 (2001). 2001-CVIM-128.
- 59) Jeong, M.-H., Kuno, Y., Shimada, N. and Shirai, Y.: Recognition of Shape-Changing Hand Gestures, *IEICE Transactions on Information and Systems*, Vol.E85-D, pp.1678-1687 (2002).
- 60) Jeong, M.-H., Kuno, Y., Shimada, N. and Shirai, Y.: Recognition of Two-Hand Gestures Using Coupled Switching Linear Model, *IEICE Transactions on Information and Systems*, Vol.E86-D, No.8, pp.1416-1425 (2003). 5.3
- 61) Johansson, G.: Visual perception of biological motion and a model for its analysis, *Perception & Psychophysics*, Vol.14, No.2, pp.201-211 (1973). 6
- 62) Johansson, G.: Visual motion perception, *Scientific American*, Vol. 232, No. 6, pp. 76-88 (1975). 6
- 63) Jovic, N., Turk, M. and Huang, T.S.: Tracking Self-Occluding Articulated Objects in Dense Disparity Maps, *International Conference on Computer Vision*, Vol. 1, pp. 123-130 (1999). 4.2.3
- 64) Ju, S.X., Black, M.J. and Yacoob, Y.: Card-board People: A Parameterized Model of Articulated Image Motion, *International Conference on Face & Gesture Recognition*, p.38 (1996). 3.4, 4.2.2, 6
- 65) Julier, S. J., Uhlmann, J. K. and Durrant-Whyte, H.F.: A new approach for filtering nonlinear systems, *the American Control Conference*, pp.1628-1632 (1995). 5.3
- 66) Kakadiaris, I. A. and Metaxas, D.: Three-Dimensional Human Body Model Acquisition from Multiple Views, *International Journal on Computer Vision*, Vol. 30, No. 3, pp. 191-218 (1998). 4.2.3
- 67) Kambhamettu, C., Goldgof, D.B., Terzopoulos, D. and Huang, T. S.: Nonrigid Motion Analysis, *Handbook of Pattern Recognition and Image Processing: Computer Vision* (Young, T.Y., ed.), Vol.2, Academic Press, pp.406-430 (1994). 2
- 68) 亀田能成, 美濃導彦, 池田克夫: シルエット画像からの関節物体の姿勢推定法, 電子情報通信学会論文誌 D-II, Vol.J79-D2, No.1, pp.26-35 (1996). 3.2, 4.2.3, 5.3
- 69) 金森 務, 片寄晴弘, 井口征士, 戸島章雄, 西山 洋: モーションキャプチャ「DigitEye3D」の実装, 電子情報通信学会論文誌 D-II, Vol.J81-D2, No.5, pp.804-809 (1998).
- 70) 加藤丈和, 深尾隆則, 羽下哲司: 対象追跡-フレーム間の類似度に着目した手法から動きのモデルに着目した手法まで-, 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研究報告, Vol.2005, No.88, pp.185-198 (2005). 2005-CVIM-150. ☆
- 71) Kehl, R., Bray, M. and Gool, L.V.: Full Body Tracking from Multiple Views Using Stochastic Sampling, *IEEE International Conference on Computer Vision and Pattern Recognition*,

- Vol.2, pp.129-136 (2005). 3.5, 4.2.3
- 72) 木本伊彦, 梶谷昭彦, 安田靖彦: スティックモデルに基づく単眼視動画像からの人体歩行運動の解析の一手法, 電子情報通信学会論文誌 D-II, Vol.J74-D2, No.3, pp.376-387 (1991). 4.2.1
- 73) Kirk, A. G., O'Brien, J. F. and Forsyth, D.A.: Skeletal Parameter Estimation from Optical Motion Capture Data, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol.2, pp.782-788 (2005).
- 74) 北嶋啓至, 萩原将文: 遺伝的アルゴリズムとファジー推論を用いた単一静止画からの人物の姿勢推定, 電子情報通信学会論文誌 D-II, Vol.J82-D2, No.2, pp.196-204 (1999). 3.6
- 75) 北島利浩, 浮田宗伯, 木戸出正継: スティックと楕円の統合モデルによる外見的特徴の少ない非剛体群の時系列位置・姿勢推定法, 電子情報通信学会論文誌 D-II, Vol.J88-D2, No.2, pp.334-346 (2005). 4.2.2
- 76) Krahnstoeber, N. and Sharma, R.: Articulated Models from Video, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol.1, pp.894-901 (2004). 3.8, 4.2.3, 4.5
- 77) Krahnstoeber, N., Yeasin, M. and Sharma, R.: Automatic acquisition and initialization of articulated models, *Machine Vision and Applications*, Vol.14, No.4, pp.218-228 (2003). 3.8, 4.2.2, 4.5
- 78) Kuch, J. J. and Huang, T. S.: Virtual Gun: A Vision Based Human Computer Interface Using the Human Hand, *IAPR Conference on Machine Vision Applications*, pp.196-199 (1994). 5.3
- 79) クラボンニューニパン, 木下宏揚, 酒井善則: スティックモデルを用いた手振りの認識, 電子情報通信学会論文誌 D-II, Vol.J77-D2, No.1, pp.51-60 (1994).
- 80) 倉掛正治, ラマカントネヴァティア: 関節のある動物体の記述と追跡, 電子情報通信学会論文誌 D-II, Vol.J76-D2, No.4, pp.854-862 (1993). 3.1, 4.2.2
- 81) Kwatra, V., Bobick, A.F. and Johnson, A.Y.: Temporal Integration of Multiple Silhouette-Based Body-Part Hypotheses, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol.2, p.758 (2001).
- 82) Lan, X. and Huttenlocher, D.P.: A Unified Spatio-Temporal Articulated Model for Tracking, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol.1, pp.722-729 (2004). 3.2, 4.2.2, 4.3, 4.5
- 83) Lan, X. and Huttenlocher, D. P.: Beyond Trees: Common-Factor Models for 2D Human Pose Recovery, *International Conference on Computer Vision*, Vol.1, pp.470-477 (2005). 3.2
- 84) Lee, M. W. and Cohen, I.: Proposal Maps Driven MCMC for Estimating Human Body Pose in Static Images, *IEEE International Conference on Computer Vision and Pattern Recognition*, pp.334-341 (2004). 3.6, 6
- 85) Leung, M.K. and Yang, Y.-H.: First Sight: A Human Body Outline Labeling System, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.17, No.4, pp.359-377 (1995). 4.2.2
- 86) Liebowitz, D. and Carlsson, S.: Uncalibrated Motion Capture Exploiting Articulated Structure Constraints, *International Journal on Computer Vision*, Vol.51, No.3, pp.171-187 (2003). 4.2.1
- 87) Lin, M. H.: Tracking Articulated Objects in Real-Time Range Image Sequences, *International Conference on Computer Vision*, Vol.1, pp.648-653 (1999). 4.2.2
- 88) Lowe, D.: Fitting Parameterized Three-Dimensional Models to Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.13, No.5, pp.441-450 (1991). 3.4, 5.2, 6
- 89) Lu, S., Metaxas, D., Samaras, D. and Oliensis, J.: Using Multiple Cues for Hand Tracking and Model Refinement, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol.2, p.443 (2003). 5.4
- 90) Magnenat-Thalmann, N., Seo, H. and Cordier, F.: Automatic Modeling of Animatable Virtual Humans - A Survey, *International Conference on 3-D Digital Imaging and Modeling*, p.2 (2003). 2
- 91) Marzani, F., Calais, E. and Legrand, L.: A 3-D marker-Free System for the Analysis of Movement Disabilities —An Application to the Legs, *IEEE Transactions on Information Technology in Biomedicine*, Vol.5, No.1, pp.18-26 (2001). 4.2.3, 4.5
- 92) Mikic, I., Trivedi, M., Hunter, E. and Cosman, P.: Articulated Body Posture Estimation from Multi-Camera Voxel Data, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol.1, p.455 (2001). 5.3
- 93) 美濃導彦, 坂口嘉之: 身体の計測, モデリングと着装シミュレーション, 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研究報告, Vol.2001, No.66 (2001). 2001-CVIM-128.
- 94) 美濃導彦: 3次元人体モデル中心処理, 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.43, No.SIG04(CVIM4) (2002).
- 95) 持丸正明, 山崎信寿: 自然な背景を含むシーンからの身体画像の抽出, 計測自動制御学会論文集, Vol.29, No.12, pp.1452-1457 (1993). 5.3
 - 96) Moeslund, T.B.: Summaries of 107 Computer Vision-Based Human Motion Capture Papers, Technical ReportLIA 99-02, Laboratory of Image Analysis, Institute of Electronic Systems, Denmark (1999). 2
 - 97) Moeslund, T.B. and Granum, E.: A Survey of Computer Vision-Based Human Motion Capture, *Computer Vision and Image Understanding*, Vol.81, No.3, pp.231-268 (2001). 2
 - 98) Moeslund, T.B. and Granum, E.: Modelling and estimating the pose of a human arm, *Machine Vision and Applications*, Vol.14, No.4, pp.237-247 (2003). 3.2
 - 99) Morasso, P. and Tagliasio, V.(eds.): *Human Movement Understanding*, Elsevier (1986). 6
 - 100) Mori, G., Ren, X., Efros, A.A. and Malik, J.: Recovering Human Body Configurations: Combining Segmentation and Recognition, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 326-333 (2004). 3.8
 - 101) Morris, D. D. and Rehg, J. M.: Singularity Analysis for Articulated Object Tracking, *IEEE International Conference on Computer Vision and Pattern Recognition*, p.289 (1998). 3.4
 - 102) Muybridge, E.: *The human figure in motion*, Chapman & Hall (1901). 6
 - 103) Muybridge, E.: *The human figure in motion*, Dover Publications Inc. (1955). 6
 - 104) 中嶋正之, 山本正信: グラフィックスとビジョンーデジタル映像処理ー, pp.72-74, オーム社 (1996).
 - 105) Nixon, M.S. and Carter, J.N.: Advances in Automatic Gait Recognition, *IEEE International Conference on Automatic Face and Gesture Recognition*, p.139 (2004).
 - 106) Nunomaki, T., Yonemoto, S., Arita, D., Taniguchi, R. and Tsuruta, N.: Multi-part Non-rigid Object Tracking Based on Time Model-Space Gradients, *International Workshop on Articulated Motion and Deformable Objects*, pp.72-82 (2000). 4.2.3, 4.5
 - 107) 大垣健一, 岩井儀雄, 谷内田正彦: 動きと形状モデルによる人物の姿勢推定, 電子情報通信学会論文誌 D-II, Vol.J82-D2, No.10, pp.1739-1749 (1999). 3.5, 4.2.1
 - 108) 大田佳人, 山際貴志, 山本正信: キーフレーム拘束を利用した単眼動画画像からの人間動作の追跡, 電子情報通信学会論文誌 D-II, Vol.J81-D2, No.9, pp.2008-2018 (1998). 3.4, 4.2.3
 - 109) Ohya, J. and Kishino, F.: Human posture estimation from multiple images using genetic algorithm, *IEEE International Conference on Pattern Recognition*, Vol.1, pp.750-753 (1994). 3.2
 - 110) 岡本浩幸, 鈴木尊人, 黒毛利学, 市原秀貢: モーションキャプチャーシステム, 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研究報告, Vol.2001, No.66 (2001). 2001-CVIM-128.
 - 111) 大垣健一, 岩井儀雄, 谷内田正彦: 動き情報と情報圧縮を用いたロバストなジェスチャ認識手法, 電子情報通信学会論文誌 D-II, Vol.J81-D2, No.9, pp.1983-1992 (1999). 5.3
 - 112) 大谷 淳, 海老原一之, 棚沢 順: 仮想変身システムー自分の姿を好みの姿に変えられるシステムを目指してー, ATR ジャーナル, No.26, pp. 6-7 (1997). 3.2
 - 113) O'Rourke, J. and Badler, N.I.: Model-based Image Analysis of Human Motion Using Constraint Propagation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.2, No.6, pp.522-536 (1980). 5.4, 6
 - 114) O'Toole, A.J., Harms, J., Snow, S.L., Hurst, D.R., Pappas, M.R., Ayyad, J.H. and Abdi, H.: A Video Database of Moving Faces and People, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.27, No.5, pp. 812-816 (2005).
 - 115) Parameswaran, V. and Chellappa, R.: View Independent Human Body Pose Estimation from a Single Perspective Image, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol.2, pp.16-22 (2004).
 - 116) Parent, R.: Computer Animation: Algorithms and Techniques - A Historical Review, *Computer Animation*, p.86 (2000).
 - 117) Pavlovic, V. I., Sharma, R. and Huang, T.S.: Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.19, No.7, pp.677-695 (1997). 2
 - 118) Pentland, A. and Horowitz, B.: Recovery of Nonrigid Motion and Structure, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.13, No.7, pp.730-742 (1991). 4.3
 - 119) Pentland, A.: Looking at People: Sensing for Ubiquitous and Wearable Computing, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.22, No.1, pp.107-119 (2000). 2
 - 120) Pentland, A. and Horowitz, B.: Recovery of Nonrigid Motion and Structure, *IEEE Transactions on Pattern Analysis and Machine In-*

- telligence, Vol.13, No.7, pp.730-742 (1991).
- 121) Plänkers, R. and Fua, P.: Articulated Soft Objects for Video-based Body Modeling, *International Conference on Computer Vision*, Vol.1, p.394 (2001). 3.5
 - 122) Plänkers, R. and Fua, P.: Articulated soft objects for multiview shape and motion capture, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.25, No.9, pp.1182-1187 (2003). 3.5, 4.2.3
 - 123) Ramanan, D. and Forsyth, D.A.: Finding and Tracking People from the Bottom Up, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol.2, p.467 (2003). 3.7, 6
 - 124) Ramanan, D., Forsyth, D.A. and Zisserman, A.: Strike a Pose: Tracking People by Finding Stylized Poses, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol.1, pp.271-278 (2005). 3.7
 - 125) Rashid, R.F.: Towards a System for the Interpretation of Moving Light Displays, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.2, No.6, pp.574-581 (1980).
 - 126) Rehg, J. M. and Kanade, T.: Model-based tracking of self-occluding articulated objects, *International Conference on Computer Vision*, p.612 (1995). 4.2.2
 - 127) Rehg, J.M. and Kanade, T.: Visual Tracking of High DOF Articulated Structures: an Application to Human Hand Tracking, *European Conference on Computer Vision*, Vol.2, pp.35-46 (1994). 3.1, 5.2
 - 128) Rehg, J. and Kanade, T.: DigitEyes: Vision-Based Human Hand Tracking, Technical Report CMU-CS-93-220, School Of Computer Science, Carnegie Mellon University (1993). 3.1
 - 129) Rehg, J. and Kanade, T.: Visual Tracking of Self-Occluding Articulated Objects, Technical Report CMU-CS-94-224, Computer Science Department, Carnegie Mellon University, Pittsburgh, PA (1994).
 - 130) Ringer, M., Drummond, T. and Lasenby, J.: Using occlusions to aid position estimation for visual motion capture, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol.2, p.464 (2001).
 - 131) Robinette, K.M. and Daanen, H.: The Caesar Project: A 3-D Surface Anthropometry Survey, *International Conference on 3-D Digital Imaging and Modeling*, p.380 (1999).
 - 132) Rohr, K.: Towards model-based recognition of human movements in image sequences, *CVGIP: Image Underst.*, Vol.59, No.1, pp.94-115 (1994).
 - 133) Rosales, R. and Sclaroff, S.: Inferring Body Pose without Tracking Body Parts, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol.2, p.2721 (2000).
 - 134) Rosales, R. and Sclaroff, S.: Algorithms for Inference in Specialized Maps for Recovering 3D Hand Pose, *IEEE International Conference on Automatic Face and Gesture Recognition*, pp.143-148 (2004). 5.5
 - 135) Rosales, R., Siddiqui, M., Alon, J. and Sclaroff, S.: Estimating 3D Body Pose using Uncalibrated Cameras, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol.1, p.821 (2001).
 - 136) Roy-Chowdhury, A. K.: A Measure of Deformability of Shapes, with Applications to Human Motion Analysis, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol.1, pp.398-404 (2005).
 - 137) Sarkar, S., Phillips, P.J., Liu, Z., Vega, I.R., Grother, P. and Bowyer, K. W.: The HumanID Gait Challenge Problem: Data Sets, Performance, and Analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.27, No.2, pp.162-177 (2005).
 - 138) 佐藤明知, 川田 聡, 大崎喜彦, 山本正信: 多視点動画画像から人間動作の追跡と再構成, *電子情報通信学会論文誌 D-II*, Vol.J80-D2, No.6, pp.1581-1589 (1997). 3.4, 3.5, 4.2.3, 5.2
 - 139) 尺長 健: 単眼視姿勢推定における基本問題の計量的可解性判定, *電子情報通信学会論文誌 D-II*, Vol.J79-D2, No.7, pp.1204-1209 (1996).
 - 140) 島田伸敬, 白井良明, 久野義徳: 確率に基づく探索と照合を用いた画像からの手指の3次元姿勢推定, *電子情報通信学会論文誌 D-II*, Vol.J79-D2, No.7 (1996). 3.2, 4.2.3, 5.3, 5.4
 - 141) 島田伸敬, 白井良明, 久野義徳, 三浦 純: 緩やかな制約知識を利用した単眼視動画画像からの関節物体の形状と姿勢の同時推定, *電子情報通信学会論文誌 D-II*, Vol.J81-D2, No.1, pp.45-53 (1998). 3.2
 - 142) Shimada, N., Shirai, Y., Kuno, Y. and Miura, J.: Hand Gesture Estimation and Model Refinement Using Monocular Camera - Ambiguity Limitation by Inequality Constraints, *International Conference on Face & Gesture Recognition*, pp.268-273 (1998). 3.2, 5.3, 5.4
 - 143) Sigal, L., Bhatia, S., Roth, S., Black, M. J. and Isard, M.: Tracking Loose-Limbed People, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol.1, pp.421-428 (2004). 4.2.3, 4.3, 6
 - 144) Sigal, L., Isard, M., Sigelman, B.H. and Black,

- M. J.: Attractive People: Assembling Loose-Limbed Models using Non-parametric Belief Propagation, *Neural Information Processing Systems* (2003). 3.1
- 145) Sminchisescu, C. and Jepson, A.: Variational Mixture Smoothing for Non-Linear Dynamical Systems, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol.2, pp.608–615 (2004).
- 146) Sminchisescu, C. and Triggs, B.: Covariance Scaled Sampling for Monocular 3D Body Tracking, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol.1, p.447 (2001).
- 147) Sminchisescu, C. and Triggs, B.: Kinematic Jump Processes For Monocular 3D Human Tracking, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol.1, p.69 (2003).
- 148) Song, Y., Goncalves, L. and Perona, P.: Un-supervised Learning of Human Motion, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.25, No.7, pp.814–827 (2003).
- 149) Stenger, B., Mendonca, P. and Cipolla, R.: Model-based hand tracking using an unscented kalman filter, *British Machine Vision Conference*, Vol.1, pp.63–72 (2001). 5.3
- 150) Stenger, B., Mendonca, P. R.S. and Cipolla, R.: Model-Based 3D Tracking of an Articulated Hand, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol.2, p. 310 (2001). 3.1, 4.2.3
- 151) Stenger, B., Thayananthan, A., Torr, P. and Cipolla, R.: Filtering using a tree-based estimator, *International Conference on Computer Vision*, Vol.2, pp.1063– 1070 (2003). 5.4
- 152) Sturman, D. J. and Zeltzer, D.: A Survey of Glove-based Input, *IEEE Computer Graphics and Applications*, Vol.14, No.1, pp.30–39 (1994). 2
- 153) 鷺見和彦, 関真規人, 波部 育: 物体検出 — 背景と検出対象のモデリング —, 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研究報告, Vol.2005, No.88, pp.79–98 (2005). 2005-CVIM-150.
- 154) 高木幹雄, 下田陽久 (編): 新編 画像解析ハンドブック, pp.323–343, 東京大学出版会 (2004). 6.5.3 身体モデルに基づく動画像解析. 2
- 155) 高橋和彦, 坂口竜己, 大谷 淳: 実時間非接触非装着型 3次元人物姿勢推定法に関する一考察, 電子情報通信学会論文誌 D-II, Vol.J83-D2, No.5, pp.1305–1314 (2000). 3.3
- 156) Tamaki, T., Yamamura, T. and Ohnishi, N.: Extraction of Human Limb Regions and Parameter Estimation based on Curl of Optical Flow, *Asian Conference on Computer Vision*, Vol.2, pp.1008–1013 (2000). 3.8
- 157) Tangelder, J. W. and Veltkamp, R. C.: A Survey of Content Based 3D Shape Retrieval Methods, *International Conference on Shape Modeling and Applications*, pp.145–156 (2004).
- 158) Tenenbaum, J.B., Silva, V.D. and Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction, *Science*, Vol.290, No.5500, pp.2319–2323 (2000). 5.5
- 159) Thayananthan, A., Stenger, B., Torr, P. H.S. and Cipolla, R.: Shape Context and Chamfer Matching in Cluttered Scenes, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol.1, p.127.
- 160) Thirion, J.-P.: Non-Rigid Matching Using Demons, *IEEE International Conference on Computer Vision and Pattern Recognition*, p. 245 (1996). 3.5
- 161) Tost, D. and Pueyo, X.: Human body animation: a survey, *The Visual Computer*, Vol.3, No.5, pp.254–264 (1988). 2
- 162) Tresadern, P. and Reid, I.: Uncalibrated and Unsynchronized Human Motion Capture: A Stereo Factorization Approach, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol.1, pp.128–134 (2004). 3.10, 6
- 163) Tresadern, P. and Reid, I.: Articulated Structure from Motion by Factorization, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol.2, pp.1110–1115 (2005). 3.10, 6
- 164) Ude, A. and Riley, M.: Prediction of Body Configurations and Appearance for Model-Based Estimation of Articulated Human Motions, *IEEE International Conference on Systems, Man and Cybernetics*, Vol.2, pp.687–691 (1999). 4.2.3
- 165) Urtasun, R., Fleet, D.J. and Fua, P.: Monocular 3-D Tracking of the Golf Swing, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol.2, pp.932–938 (2005).
- 166) Veeraraghavan, A., Roy-Chowdhury, A. K. and Chellappa, R.: Matching Shape Sequences in Video with Applications in Human Movement Analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.27, No.12, pp.1896–1909 (2005). 3.3
- 167) Wachter, S. and Nagel, H.H.: Tracking Persons in Monocular Image Sequences, *Computer Vision and Image Understanding*, Vol.74, No.3, pp.174–192 (1999). 3.1, 4.2.3

- 168) Wan, E. and van der Merwe, R.: The Unscented Kalman Filter for Nonlinear Estimation, *IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (AS-SPCC)*, pp. 153–158 (2000). 5.3
- 169) Wang, J.J. and Singh, S.: Video analysis of human dynamics—a survey, *Real-Time Imaging*, Vol.9, No.5, pp.321–346 (2003). 2
- 170) Wang, L., Hu, W. and Tan, T.: Recent developments in human motion analysis, *Pattern Recognition*, Vol.36, No.3, pp.585–601 (2003). 2
- 171) Webb, J.A. and Aggarwal, J.K.: Visually interpreting the motion of objects in space, *IEEE Computer*, Vol.14, No.8, pp.40–46 (1981). 6
- 172) Webb, J.A. and Aggarwal, J.K.: Structure from motion of rigid and jointed objects, *Artificial Intelligence*, Vol.19, No.1, pp.107–130 (1982). 6
- 173) Whittle, M.W.: Clinical gait analysis: A review, *Human Movement Science*, Vol.15, No.3, pp.369–387 (1996). 2
- 174) Wu, Y., Hua, G. and Yu, T.: Tracking Articulated Body by Dynamic Markov Network, *International Conference on Computer Vision*, Vol.2, pp.1094–1101 (2003). 4.3
- 175) Wu, Y., Lin, J. and Huang, T.S.: Analyzing and Capturing Articulated Hand Motion in Image Sequences, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.25, No.12, pp.1910–1922 (2005). 3.1, 3.2, 4.2.2, 5.3, 5.4
- 176) 山本正信, 川田 聡, 近藤拓也, 越川和忠: ロボットモデルに基づく人間動作の3次元動画追跡, 電子情報通信学会論文誌 D-II, Vol.J79-D2, No.1, pp.71–83 (1996). 3.2, 3.4, 4.2.3, 5.2
- 177) 山本正信: CV 技術による身体動作の解析, 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研究報告, Vol.2001, No.066 (2001). 2001-CVIM-128. 2
- 178) 山本正信: コンピュータビジョン技術による次世代モーションキャプチャ, 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.43, No.SIG04(CVIM4) (2002). 2, 6
- 179) 山本正信: ユビキタスモーションキャプチャ画像を使っていつでもどこでも動作計測, 画像ラボ, Vol.14, No.2, pp.1–5 (2003). 2
- 180) 山本正信: ジェスチャの計測・認識・診断技術, 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研究報告, Vol.2005, No.18, pp.59–66 (2005). 2004-CVIM-148. 2
- 181) 山本正信: ドリフト修正機能を有する動画からの身体動作推定法, 電子情報通信学会論文誌 D-II, Vol.J88-D2, No.7, pp.1153–1165 (2005). 3.4, 5.2
- 182) Yamamoto, M. and Koshikawa, K.: Human motion analysis based on a robot arm model, *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 664–665 (1991). 3.4, 4.2.3, 6
- 183) 米元 聡, 有田大作, 谷口倫一郎: 多視点動画画像処理による実時間全身モーションキャプチャシステム—視覚に基づく仮想世界とのインタラクション—, 映像情報メディア学会誌, Vol.53, No.3, pp.409–416 (2000). 4.2.1
- 184) Yu, W., Sommer, G. and Daniilidis, K.: 3D-Orientation Signatures with Conic Kernel Filtering for Multiple Motion Analysis, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol.1, p.299 (2001).
- 185) Zelnik-Manor, L. and Irani, M.: Degeneracies, Dependencies and their Implications in Multi-body and Multi-Sequence Factorizations, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol.2, p.287 (2003).
- 186) Zhang, J., Collins, R. and Liu, Y.: Representation and Matching of Articulated Shapes, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol.2, pp.342–349 (2004). 3.6, 4.4, 4.5
- 187) Zhang, X. and Liu, Y.: 3D Articulated Motion Estimation from Images, *IEEE International Conference on Computer Vision and Pattern Recognition*, pp.308–314 (2005). 3.10
- 188) Zhang, X., Liu, Y. and Huang, T.S.: Motion Analysis of Articulated Objects from Monocular Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No.4, pp.625–636 (2006). 3.10, 6
- 189) Zhao, T. and Nevatia, R.: Tracking Multiple Humans in Complex Situations, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.28, No.9, pp.1208–1221 (2004). 5.3
- 190) Zhao, T., Nevatia, R. and Lv, F.: Segmentation and Tracking of Multiple Humans in Complex Situations, *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol.2, p.194 (2001).
- 191) Zhou, H. and Huang, T.S.: Tracking Articulated Hand Motion with Eigen Dynamics Analysis, *International Conference on Computer Vision*, Vol.2, pp.1102–1109 (2003). 4.2.3