

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

山本 正 信†

動画像を使った動作計測の特徴は、身体に接することなく動作が測定できる点である。また、映像さえ残っていれば過去に遡って動作を計測することが可能である。本論文では、身体ロボットモデルをキーワードに、動画像による身体動作計測の現状や問題点を整理し、今後の課題について考えることにする。動作計測のプロセスは、次の3つのステップに分けることができる。まず、画像から身体像を抽出し、身体部位を検出する。次に、身体が多関節モデルを導入し、モデルを身体像に一致させ、姿勢を測定する。この姿勢測定法は、順運動学手法と逆運動学手法に分けられる。第3ステップは、動画像からの動作の逐次推定である。この推定は、動画像をガイドとしたロボットの動作計画と見なすことができる。各プロセスについて述べた後、さらに、得られた動作の編集による新たな動作の生成法について述べる。最後に、今後の課題として、単眼画像からの3次元姿勢推定、推定精度の評価、動画像からの身体形状推定、動作要因の推定、推定の自動化・高速化をあげておく。

Computer Vision-based Motion Capturing System

MASANOBU YAMAMOTO†

A computer vision-based motion capture has a remarkable advantage that it can measure a pose and motion of human body impermeable. This means that one can capture a human motion in movies or videos in the past. This paper surveys computer vision-based approaches for human motion capture. A process of these approaches is composed of three steps. The first step is to detect parts of human body from the image. The second step is model fitting to estimate the human pose. The model fitting can be established by direct or inverse kinematics approaches. The third step is to estimate the motion. Several constraints can uniquely determine the motion. After explaining these processes in details, techniques generating new action from captured motions are described. Finally, future works are directed, that is, a human pose from an image, an evaluation of pose estimation, a human shape from motion, guessing causes of human action and automatic capturing human motion in real time.

1. ま え が き

コンピュータビジョン(CV)の研究は、1960年代に始まった。研究対象が、積木の世界から自然界へと推移していったが、しばらくは静止したシーンが対象であった。1970年代の終わり頃から、動きのあるシーンが対象とされはじめた。動画像解析の始まりである。この頃すでに身体の動作計測が試みられていた^{15),22),23)}。しかし、実動画像を扱った研究は少なく、実動画像による身体動作計測が本格化するのには、1990年代に入ってからである。この理由として、1つは、機材の問題がある。コンピュータによる動画像の取り込みと解析が容易に行えるようになったのは最近のことである。もう1つは、身体を解析するための二-

ズが広く認識されてきたことにもよる。身体動作計測は、人間とコンピュータとのインタフェース、医療や福祉、スポーツ、放送、セキュリティ、監視、交通、アミューズメント、エンタテインメントなど多くの分野で必要とされている。

身体の動作計測は、CV以外にも、様々な分野で取り上げられてきた。特に、バイオメカニクスや人間工学では、光学式、磁気式、機械式など身体の動作を測定する様々な方式が提案されている。これらの方式に、動作をコンピュータグラフィックス(CG)上で再現する機能を加えたものを、モーションキャプチャと呼んでいる。

モーションキャプチャは、アニメーションなどの制作をはじめとして幅広く利用されているが、身体に特殊な器具を装着したり、あらかじめマーカを張り付けたりすることが多く、自然な状態で身体動作を測定することが難しい。そのため、身体に接触せずに動作が

† 新潟大学工学部

Faculty of Engineering, Niigata University

測定できる方法が望まれている。

CV 技術による動作計測は、身体に接触することなく測定できるため、次世代のモーションキャプチャと考えられる。身体の動画解析では、まず動画から特徴点を追跡し、追跡データを使って、身体領域の抽出、部位の特定と構造の推定、動作の認識などが行われる。これらは画像から次第に高次の情報抽出へと処理が進むのでボトムアップ処理と呼ばれる。これに対して、身体の運動と構造は複雑であるため、運動と構造を表すモデルを知識として用意し、トップダウン処理を行うことの必要性が認識されてきた。

身体モデルは一般に多関節モデルである。筆者ら²⁶⁾は、多関節モデルはロボットのモデルでもあることに着目し、ロボティクスの考え方を取り入れることにより、身体動作計測を進めてきた。本論文では、ロボットモデルをキーワードに、身体動作計測の現状や問題点を整理し、今後の課題について考えることにする。

次章では、身体動作計測の最初のステップとして、身体像の抽出法について述べる。次いで、3章では多関節モデルに基づく姿勢の推定法、4章では動画からの動作推定法、5章では得られた動作の編集による新たな動作の生成法についてそれぞれ述べる。最後に今後の課題を展望する。

2. 身体領域の抽出

動画をを用いた身体動作計測法には広い用途が期待される反面、問題点もある。その1つとして、一般に画像には対象としている身体以外にも様々な物体が映りこんでいることがある。そのため、まず画像から身体像の領域を分離しなくてはならない。

2.1 画像上の領域抽出

連続する画像間で濃淡値の差をとれば、動いている身体の輪郭を抽出することができる。また、背景画像があらかじめ得られているならば、画像と背景画像の差から、身体領域を抜き出すこともできる²¹⁾。ただし、これらの簡単な方法は、カメラが固定されている場合に限定されている。得られた抽出画像には、人体の輪郭や領域以外にもノイズにより様々な領域が発生し身体と区別する必要がある。小さなノイズは、あらかじめ画像に平滑化処理を施したり、差分画像に縮小と拡大を繰り返すことによって除去できる。抽出された人物像の領域を値1、背景を0とすれば、身体シルエットを表す2値画像が得られる。

2.2 立体領域の抽出

画像を使った測定法のもう1つの問題点は隠れの問

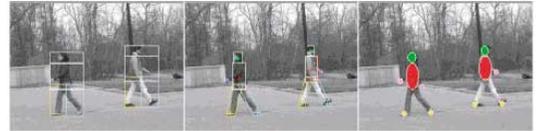


図1 身体部位の決定。左：プロポーションによる区分け，中：部位の外接長方形，右：部位の楕円近似

Fig. 1 Specification of parts of human body. Left: parts discrimination based on a human proportion, Center: circumscribed rectangles, and Right: ellipsoid model. I. Haritaoglu ©IEEE 1998

題である。隠れの問題はカメラの数を増やせばある程度解決することができる。しかし、映像間の同期やカメラキャリブレーションなどを必要とする。

複数のカメラからは、3次元情報を得ることも可能である。たとえば、ステレオカメラ対から、身体の3次元面情報を¹¹⁾、また、各カメラ視点からのビューボリュームの積から身体立体領域を得ることができる。

2.3 部位の特定

身体像が領域として抽出されたならば、次は、そこから頭や手足のような身体各部を特定する。画像がカラーであるならば、顔や手の位置は肌の色を利用して決定することができる。ただし、身体以外にも肌色領域は多く、顔や手の領域の特徴を使って区別する必要があるが、まだ万全ではない。歩行などで直立しているならば、身体のプロポーションを使っておよその位置が分かる⁴⁾。たとえば、身体の高さを1としたとき、身体像の上から1/5は頭部、首から下1/2は胴体、足元から上1/2は足部とする。そのときの各部の位置が図1に示してある。また、シルエット中の先端部分は頭や手先、足先である可能性が高い。この情報も位置の決定には有用である。一方、身体を支えるために、足または手などの体の一部は、つねに床などの環境に接しており、そこでは環境との相対速度がゼロとなる⁸⁾。したがって、逆に相対速度がゼロの部位を足先や手先と見なすことができる。

3. 姿勢の推定

身体を表すデータが、画素の集合あるいは3次元座標値の集合として得られたとする。このデータ構造は次のような欠点を持っている。

第1に、姿勢や動作を表すには、これらのデータすべてが必要というわけではなく、関節位置など要所要所のデータがあれば十分である。第2に、データが時間の経過とともにどのように推移していくかという情報が欠けている。第3に、画素データや形状データは、それぞれカメラ座標系やシーンの座標系で記述されて

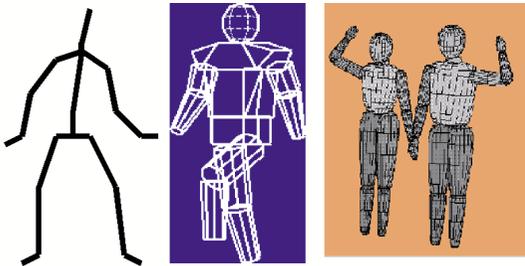


図2 身体モデル。左：スティックモデル，中：多面体モデル，右：超楕円体モデル

Fig.2 Human body models. Left: stick model, Center: polyhedrons model, and Right: superquadrics model. D.M. Gavrilu ©IEEE 1996

いることが多い。このため、データを採取した環境や使用した機材によって、同じ動作でもまったく異なったデータとして記述されてしまう。動作の比較や再現を効率良く行うためには、姿勢や動作の表現形式を統一しておく必要がある。

画像センサから得られたデータを簡潔に、しかも統一的な形式で表すために、身体モデルを次節で導入する。姿勢や動作の統一的な表現形式として、身体モデルに固有の座標系を用いる。本章では、このモデルを使った姿勢の推定法を以下に示す。

3.1 身体モデル

身体の構造を表すためのモデルとして、多関節モデルが一般的である。多関節モデルの最も簡単な例は、スティックモデルである。このモデルは、図2(左)に示されるように、スティック(棒)が端点で他のスティックとジョイント結合した構造を持ち、身体の骨格を表している。

図2(中)および(右)の多関節モデルは、身体各部位を多面体あるいは超楕円体で近似したものである。これは、スティックモデルに肉を付けたものとも考えられる。

対象となる動作に応じてモデルをさらに精密に構成する場合もある。たとえば、体のひねり運動では胴体を胸部、腹部、腰部に分ける必要があるし、背伸び運動では肩部を胴体から分離する必要がある。また、新体操やフィギュアスケートのような美しさを競う競技では手先の微妙な動きも重要である。そのためには腕の先に手や指のモデルを付け加えることも必要となる。

これらの身体モデルは、あらかじめ手作業によって作成することができる。より正確なモデルを自動的に得るためには、2.2節で述べた立体領域の抽出法やレーザレンジファインダによる身体測定などが有用であろう。

3.2 身体モデルの位置と姿勢

多関節モデルは身体各部位の階層構造で構成されている。胴体を階層構造の最上位に置き、末端に向けて頭や手、足がつながっている。階層構造内の直接の上下関係を親子関係とも呼ぶ。

この多関節モデルでは、最上位の胴体のみが並進と回転の6つの自由度を持っている。手足などの部位はそれらの親に関節でつながれているので、動きは関節位置を中心とした回転のみである。したがって、自由度は3である。図2(中)のモデルでは、全自由度は33個になる。

腕の先に手、足首の先に足、さらにその先に指がモデルに加えられるとさらに自由度は多くなる。しかし実際の姿勢を表すには、これらの自由度がすべて必要というわけではない。たとえば、指の各関節の動きは1自由度でもよいし、下腕や下脚は3自由度もいらない。上腕や上脚が2自由度あれば、1自由度で十分である。これらの場合には、自由度の総数は少なくなる。

姿勢を表す変数(パラメータ)は胴体の3次元座標値や関節角度である。これらの変数の集まりをベクトル q とする。各変数に適当な値を与えれば、多関節モデルはそれに対応する姿勢をとることができる。したがって、モデル上の点 p の座標値は、姿勢を表す変数 q の関数である。式では

$$p = f(q) \quad (1)$$

と書ける。ただし、この式では点の初期位置が省略されている。ここで、関数 f は一般に非線形ベクトル関数となる。

3.3 姿勢の推定法

身体モデルを身体像と照合し、姿勢を推定する手法を述べる。

この問題は、モデルベースドビジョン⁹⁾の延長線上にとらえることができるが、対象が身体であることへの配慮が必要である。従来、モデルベースドビジョンでは、対象が人工物であることが多く、対象のモデルを厳密に作成することができた。そのため、モデルの特徴である頂点や稜線は、画像上の特徴である特徴点やエッジにほぼ1対1に対応していた。身体はその形状や動きが複雑であるため、精密なモデル化は難しい。多関節モデルは身体の近似モデルである。身体像から観測できる多関節モデルの特徴としては、画面上への投影領域、その輪郭、手足の体軸、手先足先や頭部などの身体端点、肘や膝などの関節位置など限られている。限られた数のしかもあまり正確でない特徴どうしの対応づけということに注意すべきである。

多関節モデルは、ロボットあるいはその腕部である

マニピュレータのモデルでもある。ロボティクスでは、マニピュレータの関節角度を与えることによって指先の位置と姿勢を計算する方法を順運動学手法と呼んでいる。逆に、指先の位置と姿勢から関節角度を推定する方法を逆運動学手法と呼んでいる。身体の姿勢を推定する方法も、この順運動学と逆運動学の考え方をを使って説明することができる。

3.3.1 順運動学による姿勢推定

多関節モデルの変数に値を与えれば、順運動学手法によりモデルの位置と姿勢を計算することができる。カメラの位置とその向きが分かっているならば、得られたモデルの姿をカメラの投影面に表示させることができる。モデルの投影像を 2.1 節で得られた身体のシルエット像に重ね合わせたとき、両者が一致しているならば、モデルの位置と姿勢は実際の身体の位置と姿勢を表しているといつてよい。もし、一致していなければ、変数の値を変えて再度モデルの投影像とシルエット像を重ね合わせる。この操作を両者が一致するまで繰り返す。

一致の度合いは、たとえば、モデルの投影像とシルエット像を加えた領域を求め、それに対する重なった領域の比率によって評価できる。この比率は重なり領域が増えるに従って値が大きくなる。一方、輪郭に着目した評価法³⁾もある。モデルの輪郭上の点に対してシルエット像の輪郭上で最も近い点を求める。この2点間の距離を輪郭全体にわたって平均した距離を一致度の評価値とする。モデルが身体に近づくに従って平均距離は小さくなる。

さて、位置と姿勢の決定問題であるが、多関節モデルは姿勢の自由度が大きいので、しらみつぶしの探索は不可能に近い。局所的な探索は局所解に陥る恐れがある。そこで、大域的な探索方法として、遺伝的アルゴリズム (GA) の使用¹⁴⁾も試みられている。

一方、シルエット像の中で、頭や手足などの部位がどこにあるのか分かっているならば、部位ごとに探索を行えばよく、探索範囲は狭まる。また、求める位置と姿勢に近いものが与えられているならば、これを初期値として、局所的な探索で解を得ることができる。

このような条件は、身体を追跡するときに満たされる。なぜならば、追跡は前の時刻に置かれたモデルを次の時刻の身体像と一致させるという手順の繰返しであるので、前の時刻でのモデルが良い初期値となる。

追跡例を図3に示す。これはアルゼンチンタンゴを踊っている踊り手を4台のカメラで追跡したときの1コマである。踊り手のモデルは超楕円体で構成されている。一致の程度は、輪郭間の平均距離で評価してい

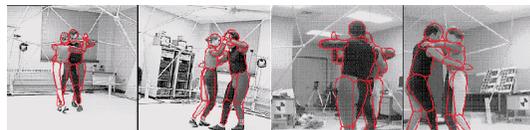


図3 多視点画像からのアルゼンチンタンゴの追跡

Fig. 3 Tracking result of dancers from multiple views.

D.M.Gavrila ©IEEE 1996

る。この追跡では、最初に胴体、次いで手、最後に足の順に照合を行っている。

追跡例は、複数のカメラからの姿勢決定法であるが、単眼カメラの場合には、対応する3次元モデルの姿勢を一意に決定できるとは限らない。1つの画像に対して複数の姿勢を候補にあげることができるときには、決定を保留しておく。動画像の前後のフレームでも対応付けを行い、動作が滑らかに推移するような姿勢列を選択する¹⁹⁾。この選択には、ビーム探索などを使って探索効率を上げることができる。

3.3.2 逆運動学による姿勢推定

モデルの手先の位置 p が指定されたとき、その位置を実現するためのモデルの位置と姿勢は、方程式 (1) を未知数 q について解けば得られる。

$$q = f^{-1}(p) \quad (2)$$

しかし、一般に複数の解が得られ、その中から正しい解を選ぶ必要がある。

身体の姿勢決定問題では、位置情報は画像から与えられる。すなわち、2章で述べたように、身体のシルエット画像を抽出し各部位が特定できれば、腕や足の体軸が画面上の直線として抽出される。肘や膝の位置は体軸の交点から得られる。また、シルエット像の先端部の中心点から、頭や手先や足先の位置が求まる。

モデル上での手先の位置を p としたとき、この投影はシルエットの手先位置に一致しなければならない。この結果、非線形方程式が1つ導かれる。肘や膝に対しても同じく非線形方程式が導かれる。また、腕や足などのモデル体軸上の点 p は、その投影位置がシルエット上の体軸に乗っていないといけない。この関係からも、非線形方程式が導かれる。

非線形方程式が多数導かれるので、これらを連立させて解を求めることになるが、解が複数個存在することもある。たとえば、上腕リンクの端点位置として、肘と肩が与えられたとき、上腕の姿勢は2通り考えられる。どちらの姿勢を選択するかは、関節角の可動範囲を考慮して絞り込むことができる^{1);2);7);20)}。このとき、姿勢の精度は肘と肩の測定精度に左右されるが、それぞれの位置を画像中に正確に見出すことは難しい。これに対し、両手両足の体軸は比較的抽出しやすく信



図4 ステレオ画像からの手の姿勢決定

Fig. 4 Pose of a hand from stereo views.

J.M. Rehg ©1994

頼度の高い特徴であるが、たかだか8本である。

モデルと身体像を対応付ける特徴が少なければ、方程式の数に対し姿勢パラメータである未知数の数が過剰になる。このような状況は、多くは単眼視画像で起こるが、多視点画像でも特徴点の数が少なければ起こる可能性がある。たとえば、頭部と手先しか位置が与えられない場合には、肘の姿勢を一意に決定することができない。冗長なパラメータには、任意の値を与えてもかまわないが、それが身体の姿勢として妥当であるとは限らない。解を一意に決定するためには、さらに制約条件を必要とする。たとえば、身体のバランスや動作の滑らかさ¹⁾や、身体パラメータの従属性²⁵⁾、などが提案されている。

逆運動学手法によって、ステレオカメラから手の指の姿勢を求めた例¹⁷⁾を図4に示す。指の体軸や先端位置が、図に重ねて示されている。この例では、姿勢は一意に決定された。

4. 動作の推定

動作の推定問題は、ロボットマニピュレータの軌道計画問題と等価である。マニピュレータの初期姿勢と最終姿勢が与えられたとき、逆運動学手法により軌道を計算することができる。ただし、身体モデルは、多自由度マニピュレータであるため、姿勢パラメータ空間内での軌道は無数に存在する。このとき、動画像の役割は軌道の教示である。

身体の姿勢推定を画像ごとに行えば、姿勢のデータ列が得られる。しかし、画像1フレームごとのモデルのあてはめには、非常に多くの計算コストを必要とする。これに対して、姿勢の変化分は動画像の変動から導かれる線形推定式を使って、容易に得ることができる^{5),26),27)}。したがって、最初のフレームでモデルのあてはめにより初期姿勢を求め、その後は姿勢の変化分を推定し、初期姿勢に累積すれば計算コストを節約することができる。

4.1 動作の逐次推定

モデルの位置および姿勢 \mathbf{q} とモデル上の点 \mathbf{p} は、式(1)で関係付けられていた。この関係は非線形であるが、モデルの位置と姿勢の変化 $\delta\mathbf{q}$ と、モデル上の

点 \mathbf{p} の変位 $\delta\mathbf{p}$ とは、ヤコビ行列を使って、次のような線形関係で結びつけることができる。

$$\delta\mathbf{p} = \mathbf{J}(\mathbf{q})\delta\mathbf{q} \quad (3)$$

ただし、 $\mathbf{J}(\mathbf{q})$ はヤコビ行列であり、 $\mathbf{J}(\mathbf{q}) = \frac{\partial \mathbf{f}}{\partial \mathbf{q}}$ である。また、 $\delta\mathbf{q}$ を動作パラメータと呼ぶ。

変位 $\delta\mathbf{p}$ は3次元移動ベクトルであり、これが与えられれば式(3)により位置と姿勢の変化を推定することができる。しかし、画像からは3次元移動ベクトルを直接与えることはできず、次のように間接的に与える。

まず、空間中の3次元移動ベクトルは、画像上では2次元移動ベクトル $\Delta\mathbf{P} = (\Delta X \ \Delta Y)$ として観測され、両者は次式で関係付けられる。

$$\Delta\mathbf{P} = \frac{1}{z} \begin{pmatrix} f & 0 & -X \\ 0 & f & -Y \end{pmatrix} \delta\mathbf{p} \quad (4)$$

ここで、 f はカメラの焦点距離である。

一方、動画像の濃淡値を $E(X, Y, t)$ と表す。移動先でも濃淡値が変化しないとすれば、画面上の移動ベクトルは次の1次方程式に束縛される。

$$E_X \Delta X + E_Y \Delta Y + E_t = 0 \quad (5)$$

ここで、 E_X, E_Y は画像の空間勾配であり、 E_t は時間勾配である。それぞれ、エッジ検出オペレータや差分オペレータにより容易に計算することができる。

式(3)、(4)、(5)から、移動ベクトルを消去すると、

$$\mathbf{G}\mathbf{J}\delta\mathbf{q} + E_t = 0 \quad (6)$$

が得られる。ここで、 $\mathbf{G} = (fE_X/z, fE_Y/z, -(XE_X + YE_Y))$ とする。モデルが身体に一致していれば、モデルから身体上の点の3次元座標値を知ることができる。したがって、上式中の z は既知となり、上式は位置・姿勢の変化分 $\delta\mathbf{q}$ を未知数とする線形推定式となる。

身体(モデル)上の多数の点からそれぞれ線形推定式を導きそれらを連立させる。その結果、線形連立方程式を得ることができる。それを次のように記す。

$$A\delta\mathbf{q} = \mathbf{b} \quad (7)$$

ただし、 A は係数行列、 \mathbf{b} は定数ベクトルである。

身体の位置と姿勢の変化は線形連立方程式(7)を解くことにより得られる。このときオプティカルフローは陽に求める必要はない。この連立方程式はカメラの視点ごとに導かれる。位置と姿勢の変化は、身体モデル固有の変数であるので、視点ごとに得られた連立方程式を、そのままさらに連立させることができる。すなわち、変化は連立方程式

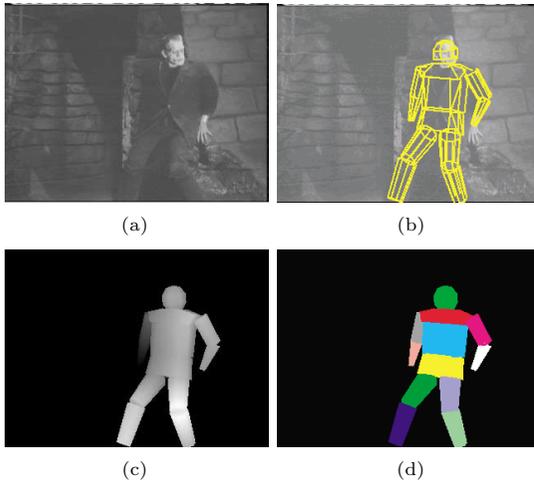


図5 モデル照合。(a) 原画像, (b) モデル照合結果, (c) 距離画像, (d) ラベル画像

Fig. 5 Model fitting. (a) the original image, (b) a overlapped wire-frame representation of the model, (c) the depth image, and (d) the region image.

$$\begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_n \end{bmatrix} \delta \mathbf{q} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_n \end{bmatrix} \quad (8)$$

を解いて得られる¹⁸⁾。ここで添え字の数字はカメラの番号を表す。このようなことが可能なのは、姿勢をモデル固有の座標系で表しているからである。

カメラの数を増やせば、多方向からの観測情報を利用できるので、隠れの解消はもとより、冗長性を利用した信頼度の高い推定が可能となる⁶⁾。

図5は、(a) 初期フレームに (b) モデルをあてはめ、モデルから得られた (c) 距離画像と (d) ラベル画像を示している。距離画像はカメラからの距離 z を濃淡値で表し、ラベル画像は部位ごとに色分けされている。ヤコビ行列は部位ごとに異なる。

このときの右大腿部の動作パラメータを推定した結果が、図6に示されている。横軸は時間軸、縦軸は回転角速度である。このように、モデルに基づく動作計測は各部位の動作パラメータを直接計算することができる。

ところで、姿勢がパラメータ空間中の特異点に陥った場合には、その後の動作を推定することが難しい。特異点を脱出するためには、姿勢を摂動させて再計算する方法¹²⁾や、リンクの伸縮を取り入れた特異点のない身体モデル¹⁰⁾も提案されている。

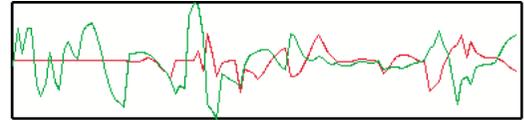


図6 運動パラメータの推定結果

Fig. 6 Captured motion curves for right-thigh. The red and green curves denote x and y components of rotation, respectively.

4.2 環境からの束縛の利用

一般に、身体はその動作環境から束縛を受けている。たとえば、スキーやスケートをしているとき、その足先は滑走面に拘束されている。また、ドアを開けるときには、手先の動きはドアノブの軌道に追従している。ロボティクスでは、環境からの束縛があらかじめ分かっているならば、この束縛を満たすように動作を計画している。

動画像から身体の動作を推定する問題では、動画像を動作決定のための束縛と見なすことができる。動画像からの束縛と環境からの束縛は、重複する場合もあるが、両者を統合することにより、冗長性を利用したノイズの抑制や動作の一意的な推定が可能となる。

環境からの拘束は、以下のように表される。身体上の点 \mathbf{p} が平面 $\mathbf{n}^\top \mathbf{x} = d$ に拘束されるとする。ここで、 \mathbf{n} 、 x 、 d はそれぞれ、面の法線、3次元座標、原点から拘束面までの距離とする。点 \mathbf{p} が $\mathbf{p} + \delta \mathbf{p}$ に移動したとき、移動先が拘束面に載るためには、

$$\mathbf{n}^\top \delta \mathbf{p} = d - \mathbf{n}^\top \mathbf{p} \quad (9)$$

であることが必要である。この $\delta \mathbf{p}$ に、式(3)を代入すれば、動作パラメータの拘束式が得られる。

身体がある軌道に沿って動いている場合には、適当な2つの面に同時に拘束されるとすればよい。さらに、空間中の1点に固定されている場合には、適当な3つの面に同時に拘束されるとする²⁴⁾。これらの環境からの拘束式と画像からの拘束式(7)(あるいは(8))を連立させて動作を決定する。

束縛は動作環境からだけではない。身体が自身の他の部位や他者と接しているときには、相手先からの束縛を受ける。この場合には多関節モデルは、相手先を含めた閉リンク系として扱う必要がある²⁰⁾。

4.3 累積誤差の解消

姿勢の差分を累積する方法は、差分の推定誤差も累積され長時間の測定では、ずれを生じる。このずれを修正する方法として、初期フレームだけでなく、最終フレームでもモデルを身体像に照合しておく。この照合によって得られた姿勢と、変化分の累積によって

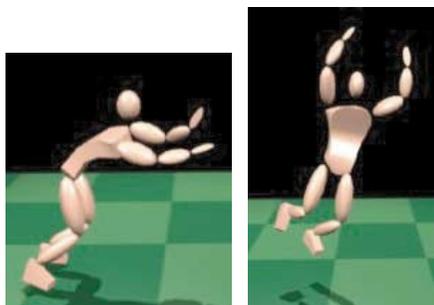


図7 左：普通の幅飛び，右：捻りを加えた幅飛び

Fig.7 Generating a twist jump (right) from a normal jump (left). Z. Popovic ©1999

得られた姿勢が最終フレームで一致するように推定値を修正するのである¹³⁾。

姿勢のずれを解消するためには、最終フレームとその1つ手前のフレームの間で、与えられた最終姿勢に達するようにモデルの姿勢を変化させればよい。これは可能ではあるが、姿勢の変化分が最後だけ異常に大きくなり、人間の動作としては不自然である。

自然な動作を得るためには、全フレームでの姿勢の変化分を再計算する必要がある。その計算は、第1に動作が滑らかに推移するように、第2に累積姿勢が最終姿勢に一致するように、第3に動画像からの拘束や環境からの拘束を満たすように行われる。この計算は最終姿勢を途中の姿勢に伝播させる繰返し演算である。ただし、動作の速さあるいは複雑さによっては、モデルの照合が最初と最終のフレームだけでは足りず、間のいくつかのフレームでも行う必要がある。

5. 動作の再生と編集

測定された動作パラメータをCGキャラクタに与えることによって、動作を忠実に再現することができる。動作を測定したときと異なる状況で再生するためには、動作データの編集を必要とするが、編集した動作が力学的な妥当性を持っていないとてはならない¹⁶⁾。図7は、通常の幅跳びの動作データを力学データに変換したうえで、ひねりを加えた幅跳びの動作を作り出している。

このことは、動作を位置や姿勢といった、見かけの記述だけではなく、動作の要因まで踏み込んで記述することの必要性を示唆している。

6. まとめと今後の課題

ロボットモデルをキーワードに、身体動作の動画像解析技術について現状を整理した。現状では、多視点カメラからは、かなり正確に姿勢や動作や身体形状を

測定することができる。しかし、ほとんどの映像は単眼カメラからの映像であるので、今後、この動作計測技術が広く使われていくためには、次のような課題に取り組む必要がある。

(1) 単眼画像からの3次元姿勢推定。1つの画像に身体モデルをあてはめるとき、得られる姿勢の確かさはモデルの正確さに依存する。カメラから観測できるのは、衣服の動きであるにもかかわらず、これまで衣服をモデルに加えることはあまりなかった。これは、衣服の形状や動きが非常に複雑でモデル化が困難であったためである。姿勢や動作を正確に測定するためには、衣服のみならず皮膚や筋肉の動きも考慮する必要がある。モデルをより精密に作れば、画像との照合により多くの計算を必要とする。大量の照合計算には、CGキャラクタを高速に生成できる最近のゲーム機の利用も一案であろう。

一方、用途によっては必ずしも精密なモデルを必要としない場合もある。この場合、画像上のあまり正確でない特徴と近似モデルとの対応付けから姿勢を推定することになる。不確実性が存在する状況で最良の推定を目指すには、データの統計的扱いが必要である。

(2) 測定精度の評価。モーションキャプチャに要求される測定精度は用途により異なる。エンタテインメントの分野では精度の高さはさほど重要ではないが、医療やスポーツの分野ではかなり重要と考えられる。測定精度を評価する基準の1つとして、3.3.1項で述べた一致の度合いを目安にすることができる。ただし、単眼画像の場合は3次元的な評価は難しい。一方、既存の接触型モーションキャプチャと比較するためには、基準となる動作データが必要である。最近話題となっている人間型ロボットを共通の測定対象とすれば、基準となる動作データが既知であり比較が可能となろう。

(3) 動画像からの身体形状測定。単眼カメラの使用では、あらかじめ作成しておいた身体モデルの利用を前提としてきたが、動作の解析と同時にモデルも作成できれば理想的である。動画像解析の初期の頃には、身体の構造と運動を同時に推定する試みがあった²³⁾。しかし、それらは適用条件が非常に厳しいものであった。非剛体の形状推定という難問に再度挑戦することは意義がある。

(4) 動作要因の推定。測定した動作を新たな動作に変形したとき、それが力学的に妥当であるためには、動作データを力やトルクなどの力学データで表しておく必要がある。そのためには、身体の質量や慣性モーメントなどをモデルに加え、動画像から直接、力やトルクを得てもよい。力やトルクは動作の要因とも考え

られる．他の要因としては，視覚や聴覚から得た情報，動作の目的や約束事などが考えられる．これらの要因と動作との関連を明らかにしておけば，要因の変化に応じた動作の生成が容易になろう．

(5) 動作や形状測定の自動化・高速化．動画画像解析技術は，既存のモーションキャプチャに比べて，優れた点が多くある．なかでも，決定的なのは，過去の映像からでも動作が測定できる点であろう．これは，映画の発明以来，現在に至るすべての映像が解析対象であることを意味している．この場合，解析のリアルタイム性は必ずしも要求されないが，解析を待っている膨大な映像データを前にしたとき，処理の自動化・高速化は必要である．実時間処理が容易になれば，人間とコンピュータとのより自然なインタラクションが実現できよう．

謝辞 図の使用を許可していただいた，D.M. Gavrilu, I. Haritaoglu, J.M. Rehg, Z. Popovic の各氏に感謝いたします．また，有益なコメントをいただきました本学玉木徹助手ならびに査読者の方々に感謝いたします．

参 考 文 献

- 1) 天谷賢治，原 裕二，青木 繁：逆解析手法による 3 次元人体運動の再構成，機械学会論文集中 (C 編)，Vol.63, No.608, pp.1167-1171 (1997).
- 2) Barron, C. and Kakadiaris, I.A.: Estimating anthropometry and pose from a single image, *IEEE CVPR00*, pp.669-676 (2000).
- 3) Gavrilu, D.M. and Davis, L.S.: 3-D model-based tracking of humans in action: A multi-view approach, *IEEE CVPR96*, pp.73-80 (1996).
- 4) Haritaoglu, I., Harwood, D. and Davis, L.S.: W^4 : Who? When? Where? What? Arealtime system for detecting and tracking people, *FG98*, pp.222-227 (1998).
- 5) 岩井儀雄，八木康史，谷内田正彦：単眼動画からの手の 3 次元運動と位置の推定，信学論，Vol.J80-D-II, No.1, pp.44-55 (1997).
- 6) Kakadiaris, I.A. and Metaxas, D.: Model-based estimation of 3D human motion with occlusion based on active multi-viewpoint selection, *IEEE CVPR96*, pp.81-87 (1996).
- 7) 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推定法，信学論，Vol.J79-D-II, No.1, pp.26-35 (1996).
- 8) Leung, M.K. and Yang, Y.H.: First sight: A human body outline labeling system, *IEEE PAMI*, Vol.17, No.4, pp.359-377 (1995).
- 9) Lowe, D.G.: Fitting parameterized three-dimensional models to images, *IEEE Trans. PAMI*, Vol.13, No.5, pp.441-450 (1991).
- 10) Morris, D.D. and Rehg, J.M.: Singularity analysis for articulated object tracking, *IEEE CVPR98*, pp.289-296 (1998).
- 11) Narayanan, O.J., Rander, P.W. and Kanade, T.: Constructing virtual worlds using dense stereo, *ICCV98*, pp.3-10 (1998).
- 12) 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，信学論，Vol.J82-D-II, No.10, pp.1739-1749 (1999).
- 13) 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からの人間動作の追跡，信学論，Vol.J81-D-II, No.9, pp.2008-2018 (1998).
- 14) 大谷 淳，岸野文郎：遺伝的アルゴリズムを用いた多眼画像からの人物の姿勢のモデルベース推定，映像情報メディア，Vol.51, No.12, pp.2107-2115 (1997).
- 15) O'Rourke, J. and Badler, N.I.: Model-based image analysis of human motion using constraint propagation, *IEEE PAMI*, Vol.2, No.6, pp.522-536 (1980).
- 16) Popovic, Z. and Witkin, A.: Physically based motion transformation, *Proc. SIGGRAPH99*, pp.11-20 (1999).
- 17) Rehg, J.M. and Kanade, T.: Visual tracking of high DOF articulated structures: An application to human hand tracking, *ECCV94*, Vol.2, pp.35-46 (1994).
- 18) 佐藤明知，川田 聡，大崎喜彦，山本正信：多視点動画からの人間動作の追跡と再構成，信学論，Vol.J80-D-II, No.6 (1997).
- 19) 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，信学論，Vol.J79-D-II, No.7, pp.1201-1217 (1996).
- 20) Taylor, C.J.: Reconstruction of articulated objects from point correspondences in a single uncalibrated image, *IEEE CVPR00*, pp.677-684 (2000).
- 21) Toyama, K., Krumm, J., Brumitt, B. and Meyers, B.: Wallflower: Principles and practice of background maintenance, *ICCV99*, pp.255-261 (1999).
- 22) Tsuji, S., Osada, M. and Yachida, M.: Tracking and segmentation of moving objects in dynamic line drawing, *IEEE PAMI*, Vol.2, No.6, pp.516-522 (1980).
- 23) Webb, J.A. and Aggarwal, J.K.: Visual interpreting the motion of objects in space, *IEEE Computer*, Vol.14, No.8, pp.40-46 (1981).
- 24) 八木下勝利，山本正信：シーン拘束を用いた人間動作の高精度動画追跡，映像情報メディア，

Vol.52, No.3, pp.331-336 (1998).

- 25) 八木下勝利, 山本正信: 固有空間を利用した単眼視画像からの人体の姿勢推定, 信学技報, PRMU99-85 (1999).
- 26) Yamamoto, M. and Koshikawa, K.: Human motion analysis based on a robot arm model, *IEEE CVPR91*, pp.664-665 (1991).
- 27) 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, 信学論, Vol.J79-D-II, No.1, pp.71-83 (1996).

(平成13年12月17日受付)

(平成14年3月8日採録)



山本 正信(正会員)

昭和26年生。昭和50年東京工業大学大学院修士課程修了。同年電子技術総合研究所(現,産業技術総合研究所)入所。平成元年~2年,カナダ国立研究協議会招聘研究員。現在,新潟大学工学部教授。工学博士。コンピュータビジョン,感性情報処理,人工現実感などの研究に従事。昭和62情報処理学会研究賞受賞。電子情報通信学会,日本VR学会,IEEE CS,ACM各会員。

(担当編集委員 中村 裕一)
