2005－CVIM－148（7）
2005／3／3

ジェスチャの計測・認識・診断技術

山　本　正　信†

映像からの動作測定法には２つの大きな利点がある。それは、身体に接触することなく自然な動作を測定することができることと、過去の人物でもその映像が残ってさえいれば、その動作を測定し再現できることである。本稿では、動作の表現法として身体多関節モデルを導入し、身体モデルの画像照合によって動作を測定する手法を述べる。さらに、動作認識の手法をパターン認識の枠組みで説明し、動作の感性情報処理について例を挙げる。

Measurement, Recognition and Surveillance of Human Gesture

Masanobu Yamamoto†

An image-based motion capturing has two major advantages. First, it can measure natural movement of human body noninvasively. Second, for even a person passed away, if his/her action is recorded on a film or video, it can reconstruct his/her movement. This paper introduces an articulated model to represent human body motion, and describes some approaches for measuring body movement by matching the model with a human body in images. Moreover, we describe various approaches for gesture recognition and an example of extraction of kansei information on action.

1. はじめに

人が最も興味を持って見る対象は人間であろう。また、人が見る動く対象のほとんども人間であろう。人は人のしぐさを見るとき、美しいとかきびきびしているなどの印象を持つことがある。また、人の動作を見て、その意味や意図を理解し、次の行動に移ったり自分と比較することもある。

動作の意味や意図あるいは受ける印象を、ビデオカメラとコンピュータによって実現しようとする試みが盛んに行われている。本論文では、映像から動作を測定し認識する手法や動きの感性情報の抽出法などについて紹介する。

映像からの動作測定法は、人間を外側から見ているために限界もあるが、２つの大きな利点がある。その一つは、身体に接触することなく自然な動作を測定することができることである。二つ目は、過去の人物でもその映像が残ってさえいれば、その動作を測定し再現できることである。これらのことから、例えば、テレビの映像から、オリンピックの選手やプロの選手の技を分析し自分自身と比較することも考えられる。また、古い映画から過去の俳優の演技を測定し、CG 映像で他の俳優と競演させることもできよう。

動作を測定する準備として、2 節では動作を表現するための身体多関節モデルについて述べる。3 節では、身体モデルと画像との照合によって動作を測定する方法を示す。4 節では、動作認識の手法をパターン認識の枠組みで説明する。5 節では、動作の感性情報処理について 1 例を挙げる。

2. 身体モデル

身体の多関節モデルの例を図 1 左に示す．このモデルは各部位が関節で繋がった木構造で表されている。各部位の連接関係を図 1 右に示す。矢印の向きは親から子への向きを表している．部位に付けられた番号は、最上位の部位である腰部を 1 とし親子順に付けられているとする。例えば、腰部、胸部、上腕、下腕の順に、部位 1，2，3，4 と名づける。

各部位は固有の座標系を持っている. 部位座標系は通常原点を関節位置（腰部や胸部では重心位置）に置き, 座標軸の一つを部位の体軸に一致させておく.

身体の位置・姿勢・運動は部位座標系の位置・姿勢・運動で表すことができる．身体各部位の姿勢は，その親の座標系を基準に表す．すなわち，部位 i の座標系 Σ_i は，その親である部位 $i-1$ の座標系 Σ_{i-1} を基準

† 新潟大学
Niigata University

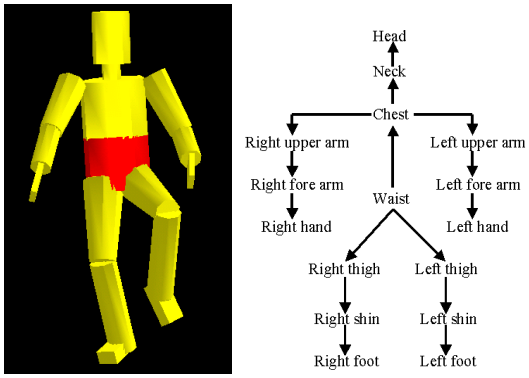

図 1　身体の多関節モデル

に並進と回転で表す。ただし、最上位の腰部の親はカメラ座標系とする。子から親座標系の変換を同次変換行列で表す。

$$T_i = \begin{pmatrix} T_i(\theta) & \mathbf{t}_i \\ \mathbf{0}^\top & 1 \end{pmatrix}$$

で表す．ここで，$T_i(\theta)$ は回転行列で，そのオイラー分解を z, y, x 軸回りの回転角 $\theta_z, \theta_y, \theta_x$ で表す。また，$\mathbf{t}_i$ は並進ベクトル $(t_{x_i}, t_{y_i}, t_{z_i})$ である。

カメラ座標系で表された点 $\mathbf{p}$ が，部位 i の座標系 Σ_i では $\mathbf{p}_i$ で表されているとする．それぞれの点の同次座標系表示は、$\tilde{\mathbf{p}}_i = (\mathbf{p}_i, 1)^\top$、$\tilde{\mathbf{p}} = (\mathbf{p}, 1)^\top$ とする。このとき部位座標系からカメラ座標系への変換は，

$$\tilde{\mathbf{p}} = T_1 T_2 \cdots T_i \tilde{\mathbf{p}}_i \qquad (1)$$

となる．

部位 i がさらに自身の座標系を基準に並進・回転運動を行ったとする。この運動を同次変換行列で

$$R_i = \begin{pmatrix} R_i(\phi) & \mathbf{r}_i \\ \mathbf{0}^\top & 1 \end{pmatrix}$$

と表す．

各部位は親の部位とリンク結合しているため、自由度は回転のみである。しかし、最上位の腰部は並進を含めた 6 自由度である。

モデルが運動したとき，位置 $\tilde{\mathbf{p}}_i$ はカメラ座標系では $\tilde{\mathbf{p}}'$ に移動したとする．この変換は，

$$\tilde{\mathbf{p}}' = T_1 R_1 T_2 R_2 \cdots T_i R_i \tilde{\mathbf{p}}_i \qquad (2)$$

で与えられる．

変換行列 T_i の位置と姿勢パラメータを変動させれば運動を表すことができる。しかし、姿勢をオイラー角で表すときには姿勢角の滑らかな変化が必ずしも滑らかな動作を表さないことがある。この動きの不連続性はジンバルロックとよばれている。ジンバルロックを避けるために、部位の動きをローカル座標系で表している。

3. 動作計測

動作の測定は、画像ごとにモデルを照合させる方法と姿勢の変化を動画像から推定する方法がある。どちらも一長一短があるが、両者を組み合わせることにより短所を補うことができる。

3.1 モデル照合

身体像が画像から抽出され各部位が特定されているとする。このとき、身体モデルを身体像に当てはめることにより身体の姿勢が推定できる。肩や肘、膝など関節位置が画像から読み取れる場合には、逆運動学手法により身体の 3 次元姿勢を得ることができる。実際、画像上で観測された部位の両端点にスティックモデルを当てはめれば姿勢が得られる。ただし、2 つの解が得られる。身体全体では、2 の部位数のべき乗の解がある。関節角の可動範囲を考慮すれば解を絞ることができる[1),2),8),20)]。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。身体のバランスや動作の滑らかさ[1)]、身体姿勢パラメータの従属性[24)] などが利用できる。

実際には画像から身体像が得られたとしても、肘や肩など関節の位置を正確に知ることは難しい。そこで、順運動学手法によりモデルの姿勢を変え、その投影像が身体像と重なったとき身体姿勢が得られる。両者が重なるまでの姿勢の探索が問題となる。

多関節モデルは自由度が大きいため、全姿勢空間の探索は GA を使うなどの工夫[13)] が必要である。通常は何らかの方法で初期姿勢を与え、局所的な探索で解を探す。身体像で部位の占める領域が分かっているならば、最初に胴体、次に胴体につながる上腕、上脚、さらに下腕、下脚の順にモデルの木構造を利用して効率よく探索することができる[4)]。

手や足が胴体など身体の他の部位と重なって見えるときは、隠れが起こりそれぞれの部位の領域を知ることは難しい。そこで、様々な 3 次元姿勢の身体サンプル画像をあらかじめ用意しておき、身体像をサンプル画像と照合させることにより姿勢を得る[15)]。サンプル画像群を PCA などでコンパクトに表しておけば、照合処理の効率化が図れる。このとき、サンプル画像の数が限られるため、必ずしも正しい姿勢が得られるわけではないが、姿勢探索の初期姿勢になりうる。

一枚の画像からでは奥行き方向の姿勢決定に曖昧さが生じる。これは、一つの身体像に対しその姿勢に複数の解釈が可能であることを意味している。このとき、

前後の画像での解釈も考慮し、動作が滑らかにつながるように[17]、あるいは姿勢の遷移確率を利用して姿勢を決定する[5]。

映像中の画像に対し順に身体モデルを照合させれば、動作を測定することができる。しかし、得られた動作は必ずしも滑らかに推移しない。これはジッター[22]とよばれ、照合に使われるサンプル画像の数が限られていることや、探索コストとのトレードオフで探索精度に限りがあることなどが原因である。

3.2 姿勢変化の累積

画像ごとのモデルの当てはめは、多くの計算コストを必要とする。これに対し、姿勢の変化分は以下に示すように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られる[7),26),27]。したがって、最初のフレームでモデル照合により初期姿勢を求め、その後は推定された姿勢の変化分を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。

モデルを身体像に一致させたとき、身体像はモデルに貼り付けられたテクスチャが画面上に投影されたものと見なすことができる。このとき，モデルの運動に応じて移動先の身体像を予測することができる．この予測画像が次の画像と一致するようにモデルの動きを求めれば，それが身体の運動となる．

モデルを運動させたとき、モデル上の点の移動先は式 (2) より与えられる。運動が小さなとき、この移動ベクトルはヤコビ行列を使って運動パラメータの線形和で表すことができる。身体に一致させたモデルから身体の 3 次元座標値が得られる。この 3 次元情報から、3 次元の移動ベクトルとその画面上への投影は線形関係となる。さらに、画面上の移動ベクトルは、画像の空間勾配（エッジベクトル）と時間勾配（画像間差分値）を使った線形式に束縛される。これらの線形関係から、最終的にモデルの運動パラメータを拘束する線型方程式が導かれる。身体像上の測定点ごとに得られるこの線型方程式を連立させれば、この連立方程式から身体の運動を推定することができる。

得られた運動を初期姿勢に加えれば、次の画像でもモデルは身体像に一致するはずである。したがって、運動の推定と姿勢への累積を繰り返せば、動作の測定が可能であるが、実際は累積を重ねるにつれモデルは身体からずれてくる。このずれをドリフトとよぶ。

3.3 モデル照合と動作累積の融合

動作の計測において、モデルの逐次照合ではジッターが問題となり、動作の累積ではドリフトが問題となる。これらの問題を解決するためには、モデル照合法と動作累積法を組み合わせるのが良い。すなわち、映像から幾つかキーフレームを選び、そこでモデル照合を行い姿勢を決定し、キーフレーム間では動作の累積を行うのである。

キーフレームでは姿勢が得られているので、中間フレームでの姿勢はロボットの動作計画法により補間することができる[21)]。この方法は動作を画像から測定する必要は無いが、キーフレーム間隔が長くなるとロボット的な動作は人特有の動作とずれてくる。したがって、動画像からの動作測定は必要である。

開始キーフレームの姿勢に測定した動作を累積した結果、それが終了キーフレームの姿勢とずれるのであるから、両者が一致するように動作を修正すればよい。動作は時間と共に滑らかに変化するとして、終了キーフレームでの姿勢を伝播させる手法が提案された[12)]。しかし、この手法は弛緩法による伝播であるため多くの計算量を必要とする。

動作は連続するフレーム間で測定することを前提としてきたが、可能ならば一つおきあるいは二つおきに測定してもよい。例えば、第 1 と第 3 フレーム間で動作が測定できたらなら、第 3 フレームでの姿勢は、初期姿勢にこの動作を加えたものと 1 − 2 と 2 − 3 フレーム間動作を累積したものとの二通りが得られる。この二通りの姿勢のうち信頼性の高い方を重視すればドリフトを抑えることが可能である[14)]。

環境からの束縛を利用してドリフトを抑えることもできる。身体はその動作環境から様々な束縛を受けている。例えば、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、ドアを開けるときは、手先の動きはドアノブの軌跡に追従している。これらの環境からの束縛を利用することにより、動作の自由度を制限しドリフトの発生を抑えることができる[25)]。

身体の動作測定ではモデルが身体からずれると測定領域内に他の運動が含まれ、それがますますドリフトを増大させる。ドリフトが一旦大きくなると、動作の多重推定や環境からの束縛を利用してもその解消は難しくなる。ドリフトを解消させる最も簡単で強力な方法は、ドリフトから 1 フレームあたりの平均ドリフトを求め、フレーム順に平均ドリフトを差し引けばよい。位置ドリフトの解消は容易であるが、姿勢ドリフトの解消は次のように行う。

一つの身体部位について、開始と終了キーフレームの姿勢をそれぞれ T_0、T_n とする。また、測定されたフレーム間動作を順に $R_1, R_2, \cdots, R_n$ とする。ここで、姿勢、動作共に直交行列で表す。動作の累積による最終姿勢は、$T_0 R_1 R_2 \cdots R_n$ であるが、ドリフトの

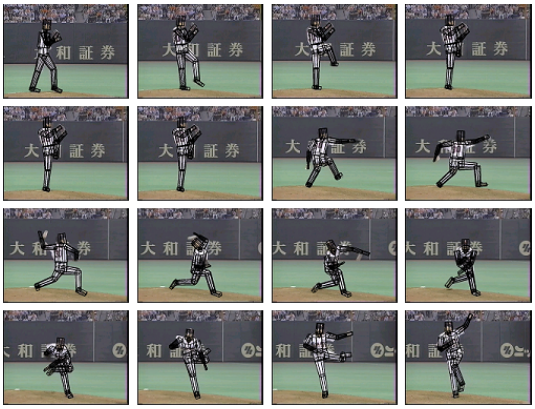

図 2　投球動作の追跡結果

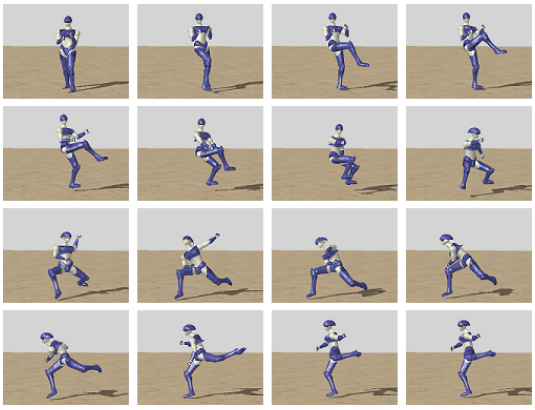

図 3　投球動作の CG 映像．追跡結果からの再構成．

ため T_n に等しくない。フレームごとの平均ドリフトの補正を X とする。この補正量は、

$$T_0 R_1 X R_2 X \cdots R_n X = T_n \qquad (3)$$

を解くことによって得られる。式 (3) は X に関する高次方程式であるが、ニュートン法による数回の逐次近似により解を得ることができる。

図 2 は野球の投球動作を追跡した結果を、モデルを画像に重ねて表示している。全部で 188 フレームである。モデルを照合させるキーフレームは 16 フレームである。図 3 はこの投球動作をバッターの視点から再現した CG 映像である。

4. 動作認識

動作を映像を通してコンピュータでどのようにして認識するかという問題について考えてみる。動作データは時系列データであるので，同じく時系列データを扱っている音声認識の手法が利用できる．一方，動作データは動作パラメータ空間中の軌道でもある。このパラメータ空間の次元を 2 次元に圧縮すれば、平面上の線画として描くこともできる．線画ならばその認識は文字認識の問題である。また、動作を基本動作のつながり、すなわち記号列で表すことができれば、動作の認識は言語理解の手法で扱える。本節では、パターン認識の手法が動作の認識にも有用であることを示す。

4.1　音声認識手法の適用

動作を認識するためには、動作例をあらかじめコンピュータに教示しておかねばならない。認識すべき未知の動作が与えられたとき、それを教示動作と比較する。動作の教示はその動作が演じられている映像そのものを用いるのが簡単である。未知の動作と教示動作をコマごと比較し、全てのコマにわたって同じであれば、その教示動作が答えである。

しかし、このコマごとの対応付けは必ずしもうまく行かない。それは、教示動作を演じた人と未知の動作を演じている人が異なれば動作も多少異なり、同じ人であっても演じた時が異なれば動作も微妙に異なってくるからである。人によって動作に速い遅いはあるし、動作中にもその速さは変化する。このような場合、未知の動作が遅い場合には教示動作の 1 コマに対し複数のコマが対応し、逆に速い場合には、未知の動作の 1 コマに教示動作の複数のコマが対応する。さらに、未知の動作の開始時刻も不明ならば、対応づけの開始もコマの順に試みなくてはならない。あらゆる対応づけの可能性について比較することになれば、全体の計算量は膨大なものとなる。

これと同じ問題は音声認識の分野でも起こっている。音声は話者によっても、同じ話者でも発声の時刻によって微妙に異なってくる。また、一連の音声信号の中から特定の言葉を発声している部分を切り出す必要もある。音声認識ではこの問題を、連続動的計画法によって解決している。基本的には，あらゆる対応付けの可能性を試す必要があるが、動的計画法はそれを効率的に行うことができる。

この動的計画法を動作認識にも適用することができる[19]。音声認識の場合との違いは、音声データは時間に関してスカラー値であるのに対して、動作データは画像であるのでベクトル値となる。そのため、比較のための計算量が多くなる。それでも、バイバイとか拍手のような簡単な動作をリアルタイムで認識できるまでになっている。

一種類の動作に対して一つの動作例をプロトタイプとして用いるとき、プロトタイプが非常に強い個性を持った人の動作であったとすれば、それとは異なる個性を持った人の動作を認識することが難しくなる。認識能力を高めるためには、様々な個性を持った人の動作例を学習しておく必要がある。

この学習の問題も音声認識の分野で扱われている。そこでは、隠れマルコフモデルが使われ大きな成果を上げている。動作の認識にも、この隠れマルコフモデルを用いることができる[28]。

学習用の動画像から個々の画像を記号化しコードブックを作成する。記号化された画像を HMM の出力とし、隠された状態の遷移確率、状態から出力への出力確率を動作例から学習する。異なる動作ごとにそれぞれの動作例から HMM の状態遷移確率や出力確率を学習し、モデルバンクに登録しておく。未知の動作に対して、その動作を最も高い確率で出力する HMM をモデルバンクから選び、それが未知の動作名となる。

4.2 文字認識手法の適用

動作の教示や学習に動画像そのものを利用してきたが、カメラの位置（視点）が異なれば動作は同じでも画像は異なってくる。したがって、異なる視点から観測した動作に対しては、改めて学習し直す必要が生じてくる。

これに対して、多関節モデルによる姿勢の表現法は、人体固有の座標系を基準にした表現であるのでカメラの視点位置に依らないデータである。

動作中の姿勢データは、パラメータ空間中に一つの軌道を描く。同じ種類の動作であればその軌道はほぼ同じ軌道を描く。一方、動作の種類が異なれば別の軌道を描くことになる。したがって、描かれた軌道の特徴から動作の識別を行うことが可能となる[3].

しかしながら、人体の姿勢を表すには多数のパラメータを必要とする。例えば、図 1 の人体モデルは 30 個のパラメータを使用している。パラメータの数が多くなれば、識別のための計算量も多くなる。また、高次元の軌道をわかりやすく表示することも難しくなる。したがって、少数のパラメータを使って、動作の特徴を失わずに表現できることが望ましい。

様々な動作の姿勢データ列を KL 展開したとき、データのばらつきが大きな順に新たな座標軸が得られる．そして，各座標軸に対応する新しいパラメータで姿勢を表すことができる．データのばらつきの大きなパラメータ値は，姿勢の変化に敏感であり独立性が高いといえる．その逆に，ばらつきが小さいパラメータ値は姿勢の変化に鈍感であり，姿勢を表すには無くても良い．これは、身体の各部は互いに関連して動いているため、独立な姿勢パラメータの数は少ないと考えられるからである。したがって，データのばらつきの大きな順に数本の座標軸を選び、これらの座標軸から構成される小さなパラメータ空間でも十分に動作の特徴は含まれている。特に、この小パラメータ空間が 2 次元の場合を固有平面と呼ぶ。

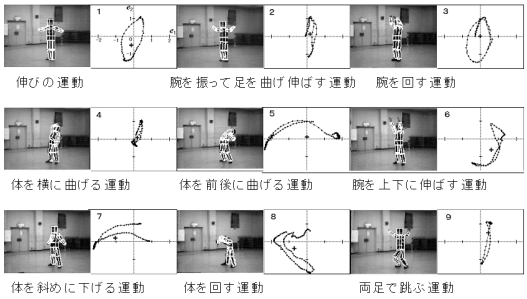

図 4 ラジオ体操の運動と固有平面上への動作軌道

固有平面上へ射影された姿勢パラメータ列も動作ごとに一定の軌道を描く。この軌道は平面上に描かれた文字とみなすこともできる。したがって、動作の認識問題は文字認識の問題と等価になる[11]。

このことをラジオ体操を例に示す。ラジオ体操第１に含まれる９種類の動作について、それらの動作中の１コマが図 4 に示されている。その右には、各動作の軌跡が固有平面上に描かれている。これらの図から、動作の種類が異なれば描く軌跡も異なることがわかる。

4.3 言語処理手法の適用

人間の行動はそのほとんどが習慣的な行動であることが多い。習慣的な行動は「一連のプログラム化された半自動的な諸行動の体系」[16] とみなすことができる。これは、行動は短い基本動作の組合せから構成されることを意味する。行動が基本動作列で表すことができれば、記号列からの動作認識は言語理解の問題と等価になる。和田ら[23] は、非決定性有限オートマトンで基本動作の遷移を記述し、この知識を使って頑強な動作認識システムを構成している。

動作の規則を文法として表すことができれば、その文法を使って動作の認識が可能である。動作の規則は観察により知ることができるが、マニュアルにより動作が定められている場合には動作マニュアルが規則集になる。動作マニュアルのある例としては、ダンスの振り付け、茶道の点前、原子力発電所などでの運転操作などがある。このうち、茶道の点前の例を示す。

茶道における点前とは、湯を沸かし、茶を点て、それを飲む動作であるが、茶道の長い歴史の中でその作法が確立されている。客や季節、使用する茶道具などに応じて様々な種類の点前が提供されている。代表的な点前は 30 種類とも言われ、それぞれ約 200 個の基本動作から構成されている。

点前は一般に図 5 のような階層構造で表すことができる。最上位は点前の種類を示し、点前は準備、喫茶、仕舞の部分動作から構成され、それぞれ具体的な

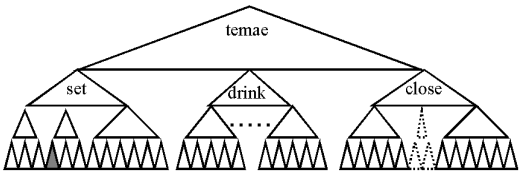

図 5 点前作法の階層構造

動作、さらには基本動作へと展開される。このような階層構造は自然言語の構造に似ており、そのため点前を記述するためには文脈自由文法が有効であると考えられる。

初歩的な点前である薄茶点前と濃茶点前の一部を、茶道の教本[18] から抜粋し表 1 に示す。濃茶点前はステップ 8 で、基本動作「主客総礼をする」（亭主が客に向かってお辞儀をする）があるが、薄茶点前では無い。その他の基本動作は共通する。

記号で「柄杓を置く」動作を h、「建水を進める」動作を k,「総礼をする」動作を r、とすれば、薄茶点前は hk、濃茶点前は hrk となる。この二つの記号列を認識（導出）する文脈自由文法を次に示す。 ここで、終端記号を $\{h, r, k\}$、非終端記号を $\{A\}$、開始記号を $\{S\}$ とする。

この文法は二つの記号列を認識することができる。では、両者を見分けるにはどうすればよいであろうか。そこで、上の表の 4 〜 5 列目に示すように、点前ごとに生成規則に確率を与え確率文脈自由文法にする。点前 hrk は生成規則 3 と 2 から導出されるが、この点前が薄茶点前である確率は、導出に使った規則の生成確率の積で与えられる。この場合は確率は 0 となる。一方、濃茶点前である確率は 1 となる。点前 hk では、これを導出する規則は 1 であるので、薄茶または濃茶である確率は 1，0 となる。以上の結果を表 3 にに示す。規則の生成確率を変えるだけで、それぞれの動作を解釈したときの確率が計算できる。 文法による動作認識では、姿勢データ列が基本動作列に変換できることが前提条件となっている。変換方法としては、基本動作の境目を動作の速度あるいは加速度の極値を使って検出する。検出された区間がどの基本動作に対応するかは、HMM や自己回帰モデルなどを使ってあらかじめ学習しておいた基本動作と比較する。

しかし、この動作の変換は必ずしも正確ではないので、誤った基本動作列からでも動作が認識できる必要がある。Ivanov と Bobick[6] は、この問題をリアルタイムパーサーを使って解決している。CYK などの従来の構文解析法は、記号列全体が与えられることが前提である。Earley-Stolcke の構文解析機は、記号列が一部ずつ与えられたとしても、予測と誤りの修正を逐次行うことができる。Ivanov と Bobick は、Earley-Stolcke 構文解析機を使ってパーキング動作のモニタリングを行っている。

しかしながら、一旦記号化した後では記号化の誤りを完全に修復させることは難しい。三富ら[10] は、動作を加速度を使って区分けした後、区間の基本動作への対応付けは保留し、対応可能なあらゆる基本動作列を残しておく。構文解析機により導出可能な基本動作列を絞り、残った候補から最も導出確率が高くなる解釈を認識結果とする。この手法はベイズの意味で最良の認識結果を与えるが、動作を区分けした区間の数が多くなると対応可能な基本動作列数が大きくなるのが欠点である。

表 1 薄茶点前と濃茶点前の動作比較

#	薄茶点前	濃茶点前
・	・・・	・・・
5	蓋置きを敷板の左隅に置く	蓋置きを風呂敷板の左隅に置く
6	柄杓を右手に持ち直す	柄杓を右手に持ち直す
7	柄杓を蓋置きの上に合をのせ引く	柄杓を蓋置きの上に合をのせ引く
8		主客総礼をする
9	左手で建水を膝前の線まで進める	左手で建水を膝前の線まで進める
10	右手で茶碗をとり膝前中央の向こうに置く	右手で茶碗をとり膝前中央の向こうに置く
・	・・・	・・・

表 2 点前の文脈自由文法

生成規則					薄茶点前	濃茶点前
1	S	→	h	k	1	0
2	S	→	A	k	0	1
3	A	→	h	r	1	1

5. 動作の評価

人は人の動作を見て、美しい、心地よい、上手だ、下手だ、きびきびしているなど様々な印象を抱く。これの印象は人間の主観にもとづいて下され、感性情報とよばれる。印象を定量化することは、なかなか難しいが、もしそれが可能なら体操競技やフィギュアスケートなどでの演技の優劣をコンピュータで判定できるかもしれない。ここでは、スキーの滑りを対象にした例を示す。

表 3 生成確率による点前の解釈

点前	薄茶点前	濃茶点前
hk	1	0
hrk	0	1

表 4 観察者の技術レベル別による判断基準（新潟大学競技スキー部による）

判断	観察者のレベル		
(指標)	初級	中級	上級
両足が同じ動き	○	○	○
上半身が安定		○	○
体軸の位置がセンター		○	○
スタンスが一定			○
重心移動がスムーズ			○
動作が途切れずなめらか			○
左右対称			○

スキーの滑りが上手あるいは下手という判定は、多少ともスキーを経験したことのある人なら下すことができる。この判定基準は，観察者の持っている専門的知識に依存する．実際，上手な滑りの基準について、ある大学のスキー部の部員に対しアンケートにより調査した。表 4 はその結果である。○印が部員により提示された基準である．部員はスキー歴の長さによって、初級、中級、上級の各クラスに分けられている。この表から、専門的知識の多い人ほど基準がより細かく，多くなっていることが分かる．これらの基準が、測定された動作の特徴と結びつけば，上手さの判定をコンピュータにより行うことが可能となる．

まずは滑走時の運動パラメータを測定しておく。ただし、出来るだけ同じ条件で、スキーの動作を比較するために、実際のスキーではなく、インラインスケートを使用した。道路上に等間隔で置かれたマーカーをスキーの旗門とみなして滑走した。カメラは１台だけで滑走方向に据え付けられている。滑走者は、初級者、中級者、上級者、最上級者の４者である。動作測定した映像の１コマを図 6 の左に示す。

スキーではターンをする際にエッジの切替が必要になる。これは角付け操作とよばれ、上脚の振り子運動で表される。上脚の付け根を中心とした左右の脚の回転角速度を図 7 に示す。この図の、左が初級者，右が上級者であり、実線が右足，点線が左足に相当している．上級者は初級者に比べて、左右の足の動きが揃っていることが分かる。これは、表 4 に提示されている基準のうち、両足の同期に相当する。この基準の良さは、左右の動きの相関値を計算することにより、数値化することができる。すなわち相関値が高いほど両足の動きが揃っていることになる。

次に動きの滑らかさを見る。これは、表 4 の「動作が途切れず滑らか」という基準に相当する。図 6 の右図は左右にターンをしているとき、横方向の胴体の速度を、初級者、中級者、上級者、最上級者について示したものである。

上級者はゆったりとした曲線を描いているのに比べて、初級者の動きは凸凹がありぎこちない。動きの滑らかさもまた数値化することができる。すなわち、理想的な動作はほぼ正弦波を描くので、動きの速度変化をフーリエ変換したとき、低周波成分の割合が多いほど滑らかな動きとみなすことができる。

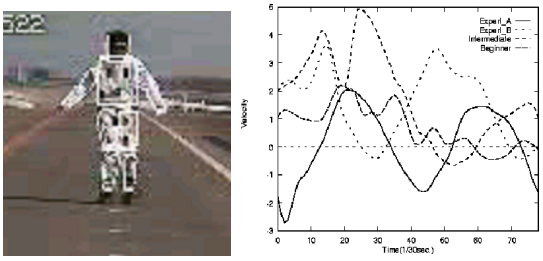

図 6 滑走動作 (左)、動作の滑らかさ (右)

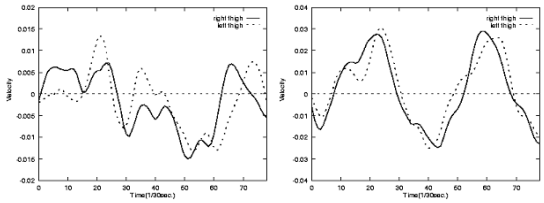

図 7 両足の同期:初心者 (左)、上級者 (右)

このようにして得られた動きの滑らかさを表す値と，先にあげた両足の同期を表す評価値を加え合わせて総合的な評価を行ってみると、上級者になるほど評価値が高くなった．これは人間の目でみたときの技量の評価と一致した．

6. おわりに

動作の測定では、キーフレームでのモデル照合とキーフレーム間での動作追跡を組み合わせることが、現時点での最良の手法であろう。しかし、実際に両者を組み合わせた例は少ない。これは、画像から身体を抽出し部位を特定することが、画像認識の基本問題であり、いまだ汎用手法が存在しないことがネックとなっている。問題を簡単にするためには、対象とするシーンや動作を限ることが必要である。

一方、動作の認識では、演じる人の動作を中心に議論してきたが、人が使用する道具や対象に着目することも考えられる。茶道では茶筅や茶碗の動き、料理では包丁の動きなどからでも動作を認識することが可能である。また、ダンスでは音楽や音声の情報も重要である。ただし、どの情報チャンネルでもそれだけで十分であるとはいえない。原子力発電プラントでは、運転員と計器や操作卓との関係が重要なように、マルチモーダルな情報を如何に関係付け如何に相補えるか[9)]

がこれからの動作認識の鍵と思われる。

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 (C編),vol.63. no.608, pp.1167-1171, 1997
2) C.Barron and I.A.Kakadiaris, “Estimating anthropometry and pose from a single image”, IEEE CVPR00, pp.669-676, 2000
3) L.W.Campbell and A.F.Bobick, “Recognition of human body using phase space constraints”, Proc. ICCV95, pp.624–630, 1995
4) D.M.Gavrila and L.S.Davis, “3-D model-based tracking of humans in action: a multi-view approach”, IEEE CVPR96, pp.73–80, 1996.
5) 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”, 信学論, Vol.J85-D-II, No.8, pp.1291-1299, 2002.
6) Y.A. Ivanov and A.F. Bobick, “Recognition of visual activities and interactions by stochastic parsing”, IEEE PAMI, Vol.22, No8, pp.852–872, 2000
7) 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”, 信学論,vol.J80-D-II, no.1, pp.44-55, 1997
8) 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
9) 川島宏彰、松山隆司、”連続状態モデル間の相互作用に基づく多視点動作認識 ”、信学論,vol.J82-D-II, no.12, pp.1801–1812, 2002
10) 三富文和、藤原冬樹、山本正信、佐藤泰介、“習慣的な行動の確率文脈自由文法に基づくベイズ推定”, 信学論, 2005 （印刷中）
11) 大野宏、山本正信，“文字認識手法を用いた固有平面上での動作認識”、情処論、Vol.40, No.8, pp.3134–3142, 1999
12) 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像からの人間動作の追跡”, 信学論,Vol.J81-D-II, No.9, pp.2008–2018, 1998
13) 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
14) A.Rahimi, L.-P.Morency and T.Darrell, “Reduction drift in parametric motion tracking”, ICCV’01, pp.315-322, 2001
15) A.A.Rfros, A.C.Berg, G.Mori and J.Malik, “Recognizing action at a distance”, Proc. ICCV03, pp.726-733, 2003
16) 塩沢由典、”複雑さの帰結 ”、NTT 出版、1997
17) 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.
18) 千宋室、”裏千家茶道教科　点前編 全 17 巻”, 淡交社、1976
19) 高橋勝彦，関進，小島浩，岡隆一, “ジェスチャー動画像のスポッティング認識”，信学論，Vol.J77-D-II,No.8,pp.1552–1561, (Aug.,1994)
20) C.J.Taylor, ”Reconstruction of articulated objects from point correspondences in a single uncalibrated image”, IEEE CVPR00, pp.677-684, 2000
21) C.Tomasi, S.Patov and A.Sastry, “3D tracking = Classification + Interpolation”, Proc. ICCV03, pp.1441-1448, 2003
22) L.Vacchetti, V.Lepetit and P.Fua, “Fusing online and offline infotmation for stabel 3D tracking in real-time”, CVPR’03, pp.II241-248, 2003
23) 和田俊和、佐藤正行、松山隆司、”選択的注視に基づく複数対象の動作認識 ”、信学論,Vol.J82-D-II, No.6, pp.1031–1041, 1999
24) 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
25) 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像追跡”, 映像情報メディア,Vol.52, No.3, pp.331–336, 1998
26) M.Yamamoto, K.Koshikawa, “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次元動画像追跡”, 信学論,Vol.J79-D-II, No.1, pp.71–83, 1996.
28) 大和淳司，大谷淳，石井健一郎、“隠れマルコフモデルを用いた動画像からの人物の行動認識”，信学論、Vol.J76-D-II,No.12,pp.2556–2563, 1993